Computing in Civil Engineering

Proceedings of the 2013 ASCE International
Workshop on Computing in Civil Engineering

Edited by   Ioannis Brilakis, Ph.D., SangHyun Lee, Ph.D.,
and Burcin Becerik-Gerber, DDes

ASCE

COMPUTING IN CIVIL ENGINEERING

PROCEEDINGS OF THE 2013 ASCE INTERNATIONAL
WORKSHOP ON COMPUTING IN CIVIL ENGINEERING

June 23–25, 2013
Los Angeles, California

SPONSORED BY
Technical Council on Computing and Information Technology
of the American Society of Civil Engineers

EDITED BY
Ioannis Brilakis, Ph.D.
SangHyun Lee, Ph.D.
Burcin Becerik-Gerber, DDes.

Preface

The 2013 ASCE International Workshop on Computing in Civil Engineering (IWCCE) was held in Los Angeles from June 23–25, 2013. The IWCCE was created by the ASCE Technical Council on Computing and Information Technology (TCCIT) to advance professional knowledge and improve the practice of civil engineering by encouraging the effective use of current and emerging computing and information technologies. The TCCIT promotes computing research and education across all aspects of civil engineering sub-disciplines. The IWCCE has been held in several states across the United States, and has produced a diverse spectrum of highly cited papers that have been published in proceedings by the ASCE and that are available from the ASCE database. The IWCCE is held bi-annually and always produces a compelling illustration of how recent advances in data sensing, information analysis (visualization, modeling, and simulation), and knowledge discovery address a number of challenges in civil engineering.

Co-sponsored by Kiewit Construction Inc., IWCCE 2013 offered something very unique by broadening the spectrum of civil engineering areas represented; it did so through encouraging the submission of papers from all civil engineering sub-disciplines. This diversified the participation of experts working in computing research in civil engineering. The result was that over 100 participants attended from 16 countries and 3 continents.

Preparation for the workshop began in 2011 along the lines of previous workshops and under the direction of the TCCIT. The major change included a revised paper review process that focused on adopting a tiered scale for accepted papers to ensure that the highest visibility is given to the best papers, while limiting the number of oral presentations to 3 concurrent/parallel sessions, and retaining additional papers through poster format presentations. Other major changes included having a dedicated local chair for smooth workshop delivery, and actively including all civil engineering areas in all aspects of the event to promote scientific diversity.

As a result, the workshop received 240 abstracts and eventually accepted 112 papers covering the following 5 state-of-the-art computing disciplines in civil engineering: Resilience and Smart Structures, Sustainability and Environment, Geo-Modeling & Geomechanics, Intelligent Transportation, and Construction Automation and Information Technology. These papers were the result of rigorous peer review processes for both the abstracts and papers, with the assignment of at least two reviewers for each abstract and paper. Only outstanding papers have been included in the proceedings.

The workshop commenced officially with a co-located poster session and a welcome reception on June 23rd, during which 22 papers were presented via the poster format. This was followed by 2 days of keynote, general, and parallel sessions, and by TCCIT committees'

meetings. The keynote talks were delivered by the following leaders in industry and academia: Professor Chimay J. Anumba who talked about "Intelligent Context-Aware Systems for Design and Construction"; Professor Campbell R. Middleton who talked about "Navier's Straightjacket – The Legacy, Limitations and Opportunities"; and Mr. Tom Harter, Mr. Jay Vose and Dr. Thomas Liebich who talked about "Open Data Exchange Standards for Infrastructure Opens the Door for Enterprise Resource Planning Systems Integration". All of this culminated at the banquet with a keynote given by Professor Feniosky Peña-Mora who talked about "nD BIM+AR: n-Dimensional Building Information and Augmented Reality Models for Automated Construction Performance Monitoring". 18 excellent papers were presented in general sessions, followed by 72 papers in parallel sessions. At the banquet, 5 area awards and 1 overall award were given for the best papers in the conference. The TCCIT thanked the conference organizing committee for their excellent work.

All the recognition, though, does not belong to us alone. We would like to thank the Sonny Astani Department of Civil and Environmental Engineering at the University of Southern California for their support and infrastructure. The ASCE TCCIT provided a variety of organizational assistance and their membership formed a large part of the reviewer pool for the workshop. Kiewit Corporation sponsored one of the workshop luncheons and provided excellent keynote speakers. The European Group for Intelligent Computing in Engineering contributed additional sessions and several of their members were a significant part of the reviewer pool. We look forward to continued close cooperation between these communities. Finally, the following students also merit special mention for their tireless work in assisting the organizational committee: JoonOh Seo at the University of Michigan, Stefania Radopoulou at the University of Cambridge, and Saba Khashe at the University of Southern California.

We hope you enjoyed the quality technical sessions at the workshop and that your stay in Los Angeles was outstanding!

Ioannis Brilakis, Ph.D., University of Cambridge, Chair
SangHyun Lee, Ph.D., University of Michigan, Technical Chair
Burcin Becerik-Gerber, DDes., University of Southern California, Local Chair

2013 ASCE International Workshop on Computing in Civil Engineering

Acknowledgments

We would like to thank the area track chairs and technical committee members for their contributions to the paper review and selection process. We would also like to thank the advisory committee, who provided feedback throughout the planning and execution of the event.

1. Track Chairs

- Resilience & Smart Structures: Jerome Lynch, Erik Johnson
- Sustainability & Environment: Kumar Mahinthakumar, George Ban-Weiss
- Geo-Modeling & Geomechanics: Dominic Assimaki, Jose Andrade
- Intelligent Transportation Jorge Laval, Wenlong Jin
- Construction Automation & IT (Data Sensing & Analysis): Nora El-Gohary, Amr Kandil,
- Construction Automation & IT (Visualization and Simulation): Fernanda Leite, Yong Cho
- Construction Automation & IT (Information Modeling): Timo Hartmann

2. Advisory Committee

- Burcu Akinci
- Carlos Caldas
- Ian Flood
- Renate Fruchter
- James H. Garrett
- Raymond Issa
- Hani Melhem
- J.P. Mohsen
- William O'Brien
- Feniosky Peña-Mora
- Ian Smith
- Lucio Soibelman
- Yimin Zhu

3. Technical Committee

- Amir AghaKouchak
- Changbum Ahn
- Saurabh Amin
- Mario Berges
- André Borrmann
- Albert Chen
- Fei Dai
- Mani Golparvar-Fard
- Jie Gong
- Javier Irizarry
- Byungil Kim
- Christian Koch
- Jitendra Kumar
- Tarek Mahfouz
- Carol Menassa
- Ram Rajagopal
- Kay Smarsly
- Mohsin Khalid Siddiqui
- Pingbo Tang
- Pieter de Wilde
- Zhenhua Zhu

Contents

Resilience and Smart Structures

vii

Intelligent Transportation

Construction Automation and IT: Information Modeling

xii

Author List

A Stochastic Finite Element Approach to Determine the Safety of Suspension Bridge Cables

A. Montoya[1], R. Betti[2,3], G. Deodatis[2,4], and H. Waisman[2,5]

[1]University of Texas at San Antonio, Department of Civil and Environmental Engineering, One UTSA Circle, San Antonio, TX 78249; PH (210) 458-7516; FAX (210) 458-6475; email: arturo.montoya@utsa.edu
[2]Columbia University, Department of Civil Engineering and Engineering Mechanics, 500 West 120th Street #610, New York, NY 10027; PH (212) 854-3143; FAX (212) 854-6267; email: [3]betti@.civil.columbia.edu, [4]deodatis@civil.columbia.edu, [5]waisman@civil.columbia.edu

ABSTRACT

A new methodology to determine the safety of suspension bridge main cables is proposed and illustrated on a corrosion-deteriorated cable composed of 9061 wires. The approach is the first one incorporating a finite element (FE) model to predict the cable's failure load, accounting for load recovery due to friction in broken wires and simulating the reduced cable's strength as a three dimensional random field. In order to obtain the breaking load of a cable, the load is increased gradually (quasi-static loading) in a cable's FE model, having wires break a few at a time according to their residual strength. Because of the load transfer to surrounding wires, the breakage of an individual wire affects the stress state inside the surrounding wires. This local damage eventually causes a global reduction in the load carrying capacity of the cable, up to a complete failure. The safety of the cable is determined through a Monte Carlo simulation, in which the reduced strength of the cable is generated for every realization through the Spectral Representation Method (SRM) and is input as a material parameter in the FE model. The statistics of the load that will drive a suspension bridge cable to failure under a hypothetical deterioration state are obtained at the end of the simulation.

INTRODUCTION

The structural function of the main cables in suspension bridges is to transfer the tension load, derived by supporting the roadway, to the towers. The main cables are composed of thousands of high strength parallel steel wires with a diameter of approximately 5 mm bundled together in strands either built in situ or prefabricated. These strands are then compacted and tightened together and eventually the cross section of the cable becomes semi-circular. The wires in pristine conditions have a strength ranging from 1570 MPa to 1800 MPa. However with aging, fatigue loading, and harsh environmental conditions, the wires strength reduces significantly (Shi et al., 2007). Field observations of aging suspension bridges indicate serious distress of

cable wires. Many wires are corroded and even broken, jeopardizing the safety of the entire bridge. Moreover field surveys performed by Suzumura and Nakumura (2004) have indicated that the corrosion level varies according to the wire positions inside the cable. Current inspection procedures rely on visual observations which are considered unreliable and expensive. Today the biggest challenge suspension bridge owner authorities face is to estimate the current and remaining safety of a main cable in order to help them decide whether it is necessary to provide immediate maintenance or rehabilitation (or even replacement) of such cables.

In order to provide an accurate assessment of the cable's remaining strength, it is critical to model the load transfer mechanism between wires and the corrosion uncertainty within the cable. One of the most important structural characteristics of suspension bridges is the high internal redundancy of its main cables. If a wire breaks, the load carried by that wire is redistributed to the unbroken wires: Montoya et al. (2012) showed through some numerical examples that, in the vicinity of the break, the neighboring wires take most of the load burden while, at some distance away from the break, the broken wire recovers fully or partially its load carrying capacity due to friction. The overall carrying capacity of the cable decreases as individual wires start to break and, as the number of broken wires increases, it might reach a point that corresponds to the complete failure of the cable.

The objective of this paper is to introduce a methodology that determines the safety of an existing cable by estimating the failure load of a suspension bridge main cable. This is accomplished through the Monte Carlo Simulation by generating a number of realizations of the cable's strength within a FE model.

PROPOSED METHOD TO DETERMINE THE CABLE'S SAFETY

The Finite Element Method can be used as the means to incorporate the interaction mechanisms between wires into the assessment of the remaining strength of a cable. A FE model can serves as an invaluable tool for evaluating bridges that exhibit distress signs. In addition, because of the uncertainties associated with corrosion, a stochastic analysis is required to determine the effects of corrosion uncertainty on the cable's failure load. The drawbacks associated with generating a model for a large number of wires and integrating it into a stochastic analysis is the computational effort associated with it. However, nowadays parallel computing provides a feasible way to solve stochastic mechanics problems.

The approach proposed in this paper estimates the failure load of a suspension bridge main cable by performing a finite element analysis coupled with a Monte Carlo Simulation. Each realization of the simulation considers different deterioration conditions in each wire. The cable considered for this study has a diameter of 24.37 cm, a prescribed length of 18.3 m, and is composed of 9061 wires with a diameter of 0.4826 cm as shown in Figure 1. The length of the model is associated with the recovery length of a broken wire within a cable. Assuming the cable has been well compacted, the overall strength of the cable can be determined from a segment of limited length, since good compaction of the wires guarantees load recovery in broken wires (Mateo, 1994). Three clamps having a width of 20 cm are placed over the length, spaced at 6.096 m from each other (see Figure 1b). The cable is pulled at both ends (in the length direction: z-axis,).
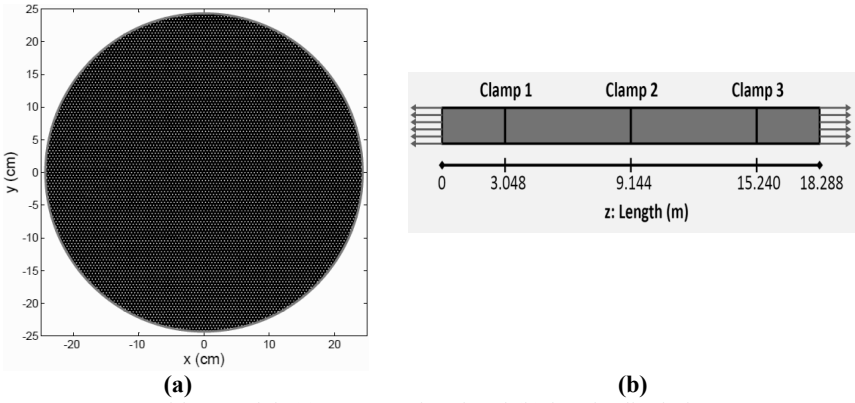
(a) **(b)**
Figure 1. Cable's model: (a) cross sectional and (b) longitudinal views.

Finite Element Model for Main Cables
In order to facilitate a full stochastic analysis of the problem and reduce the resulting computational effort to manageable levels, truss elements were selected to model the wires. Figure 2(a) illustrates a segment of two wires in a three dimensional space, exemplifying the behavior between a broken and an unbroken wire. The left wire (simulating an unbroken wire) carries an axial traction T_z at the end. The wires transfer stresses through friction related shear forces that are generated along their contact line of interaction. This interaction allows the broken wire to recover load and alleviates the load burden on the unbroken wire. An equivalent system is presented in Figure 2(b) when modeling wires with truss elements. Truss elements having a stiffness equivalent to the wire are placed at the center of each wire. The external force resulting from the axial traction is placed at node 1 of the unbroken wire. The frictional interaction between wires can be modeled using spring elements connecting both wires. The shear stress at the contact line is translated into a shear force, $F_\tau(z)$, acting at the nodes of the spring elements.

Coulomb's Friction Law (e.g. Wriggers (2006)), specifies that the maximum shear force, F_τ, that can be transferred between two surfaces is a function of a normal force:

$$F_\tau \leq \mu F_N$$

where μ is the friction coefficient (an empirical property of the contacting materials), F_N is the normal applied force, and F_τ is the resulting tangential force due to friction in a direction that is opposite to the motion of the object. For main cables, the normal force is provided by the compaction action of the steel bands. The pressure generated due to a concentrated force decays to zero exponentially when moving in the longitudinal z-direction as specified by the Boussinesq's solution to a point load applied on a half space of a semi-infinite solid (Boussinesq, 1885). Thus, the clamping action and frictional interaction between wires takes place in very close proximity to the clamps.

(a) **(b)**
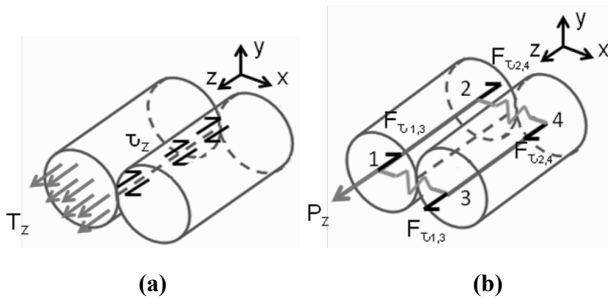
Figure 2. (a) Load transfer mechanism between a broken and an unbroken wire and (b) an equivalent finite element model

An equivalent user friction model to capture the Coulomb's Law consists on providing an elasto-plastic behavior to the spring elements between the wires in contact. This concept was proposed by Waisman et al. (2011). The stick behavior is characterized by the linear elastic stiffness and the slip part by the perfect plasticity, where the yielding force F_T represents the maximum shear force that can be transferred at that point. The yielding force decays when moving away from the clamp as observed in Figure 3(a) for four different locations from the clamp. For illustration purposes, the mesh for 19 parallel wires is shown in Figure 3(b). Note that the spring elements connect wires in the x-y plane and provide frictional interaction in the longitudinal direction.

In order to obtain the breaking load of a cable, the applied load is increased incrementally (quasi-static loading), having only a few wires break at a time. For breakage to occur, the stress has to exceed the Reduced Tensile Strength (RTS) of the element. The RTS is input as a material property of the element before loading begins and is obtained according to the Spectral Representation Method (see next section). At the initial step of the analysis, all wires are assumed to be unbroken; thus, the spring elements are inactive. As the cable is loaded, the sections of the wires that reach their RTS will break. This break is reflected in the model by eliminating the stiffness of the individual wire element, the so called element deletion technique. As the element breaks, friction on the surfaces between the broken wire and the surrounding one transfers load between adjacent wires. Numerically, this is simulated by activating user friction model elements (springs) connecting the sliding wires. Montoya et al. (2012) emphasize that when such a break happens in one wire, the load in other wires (depending on their proximity to the broken wire) will automatically be adjusted and hence may enter a critical stress state which will lead to their failure, i.e. the domino effect.

Modeling StrengthVariation in Main Cables

Field inspections have indicated that in small regions within the cable's cross section, neighboring wires will be at a similar corrosion stage. However, current methodologies treat the wire strengths as uncorrelated (Mateo et al., 1994) or only account for the correlation along its length (Shi et al., 2007). The weakest points of the simulated strength of two adjacent wires can be considerably far from each other.

This implies that the one dimensional, decoupled random field approach cannot simulate homogeneous-like corrosion in small regions within the cross section of the cable.

(a)   (b)
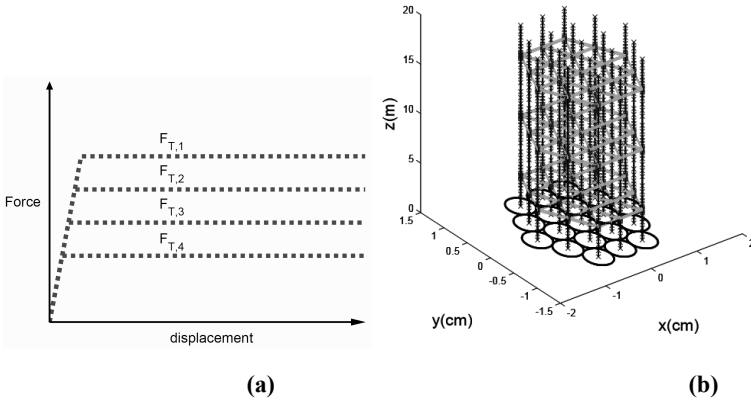
Figure 3. (a) Elasto-plastic behavior of the spring (green) elements to account for friction induced by the clamps (b) and cable's FE model.

The strength of a corrosion-deteriorated cable is simulated as a three-dimensional stochastic field by the Spectral Representation Method (SRM) (Shinozuka and Jan, 1972). This approach considers the spatial correlation of the cable strength both within the cross-section (over different wires) and along the length of the cable. Shinozuka and Deodatis (1996) have provided the theoretical background for simulating a homogeneous stochastic field in three dimensions. The prescribed three dimensional Spectral Density Function (SDF) required by the SRM formulation can be obtained by combining a SDF that captures the behavior in the cross section of the cable (x-y plane) and a SDF that accounts for the spatial correlation along the length (z-direction). The correlation is obtained from field data obtained from an actual main cable. Sample wires extracted from the cable are cut into small specimens of appropriate length (usually 25 – 30 cm) for tensile testing. A standard tensile strength test is performed on each sample to define the average strength and average elongation and their statistical deviations.

The pool of data for this study came from a thorough inspection performed on the Williamsburg Bridge cable in 1988 (Steinman,1988). The data indicated a strong dependency of the strength with the y-coordinate (cable's top to bottom). The computational time increases considerably when generating sample functions of a stochastic field in three dimensions according to the SRM. However, the Fast Fourier Technique (FFT) decreases the computational time dramatically, facilitating a three-dimensional stochastic analysis of the cable's strength (Brigham, 1988). Figure 4 shows the strength generated for three pairs of adjacent wires according to the FFT technique for three-dimensional stochastic fields (Shinozuka and Deodatis, 1996). The wire pairs are taken from the bottom, center, and top sectors of the cable's cross

section. The distribution of the wire strength corresponds well with the conditions observed in the visual inspections of the cable.
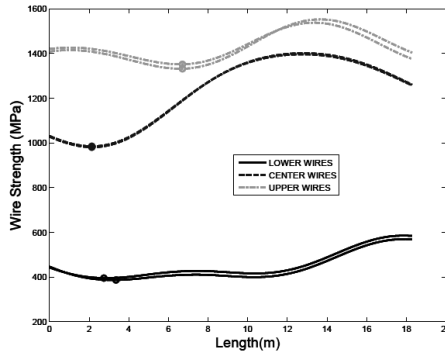
Figure 4. Generated strength of wires by considering the spatial correlation of the cable strength in space. The weakest point along the length is represented by **o**.

PARALLEL COMPUTING

Due to the overall number of equations of the model, parallelization of the problem is essential to facilitate the stochastic analysis that determines the effect of corrosion-related uncertainties on the cable's failure load. The mesh, composed of truss and spring elements, is independent of the realization; while the material property that determines the wire breaks (RTS) in each wire changes for each run according to the stochastic representation. The simulation methodology is illustrated in the flow chart of Algorithm 1.

Algorithm 1
1. Read Input Files. Store general finite element information.
2. **Loop**[1] over each Monte Carlo realization.
    - A. Set the displacement vector and force vector to zero. **u=0, F=0.**
    - B. Generate the wire corroded strengths
        - a. **Loop**[2] over each displacement increment (20 steps through the displacement controlled method).
            - i. **Loop**[3] for convergence of Nonlinear Problem
                - 1. Solve Linear Equations
            - ii. **End Loop**[3]
            - iii. Increase displacement of the cable , $u_{k+1} = u_k + \Delta u$
        - b. **End Loop**[2]
3. **End Loop**[1]

RESULTS

The failure load of the cable is obtained through a Monte Carlo Simulation (MCS) with 1000 realizations. The cable is loaded using the displacement controlled approach for twenty steps at increments of $\Delta u=0.2298$ cm. The clamps are assumed

to be sufficiently tightened to provide full recovery after one cable band. Figures 5 shows a sample step for a run of the model. The force-displacement plot (Figure 5a) compares the response of the cable to the response of a cable in perfect conditions. The cross-sectional plot (Figure 5b) is at the critical section along the cable's length where the greatest number of broken wires is found for this run. A cluster of broken wires (dark blue wires indicating zero stress) with random pattern is formed at the bottom sector. A highly stressed (orange section) region surrounds this cluster of broken wires.

(a)   (b)
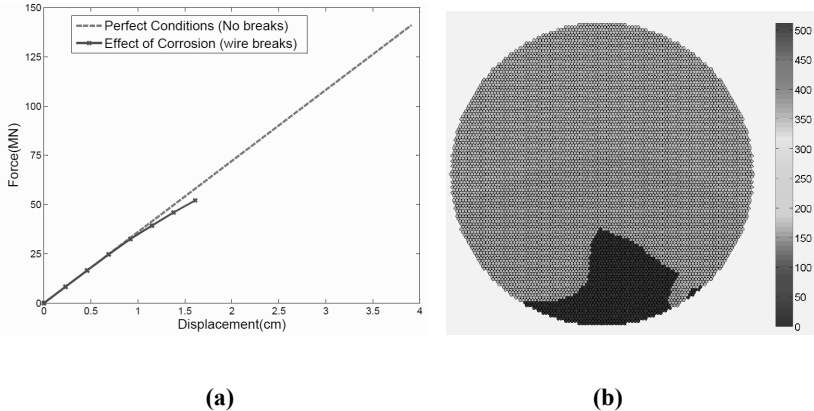
Figure 5. Sample Step of a Realization: (a) Force-Displacement Curve of the cable and (b) Longitudinal Stresses (MPa) in the 9061 wires.

Load redistribution allows the cable to carry the new load level with an uneven distribution among wires. Eventually, the cable fails in a domino like pattern in which the failure of weak wires causes other less weak wires to fail at the same external load. The load is kept constant for each of the 1000 realizations, but the cable fails at a different load as shown in Figure 6(a). The response of the system is nonlinear for each realization; the stiffness of the overall cable significantly decreases after the breaks. The corrosion-related uncertainties play a major effect on the breaking load of the cable, as significant scattering is observed in the failure load. The histogram for the breaking load is illustrated in Figure 6(b).

CONCLUSIONS

Reliable assessment of the cable's remaining capacity is needed by bridge authorities in order to make important decisions regarding maintenance and rehabilitation of main cables in suspension bridges. The proposed methodology provides an estimation of the distribution of the failure load, which may help preventing an unexpected failure. This approach is the first one incorporating a finite element analysis within a Monte Carlo simulation to evaluate the performance of main cables. This method accounts for the correlation of the cable's strength along its length and within the cross section. The proposed formulation is capable of capturing the nonlinear behavior of the cable's strength as wires break, starting from the most corroded areas in the cable's cross-section (usually the bottom sector).
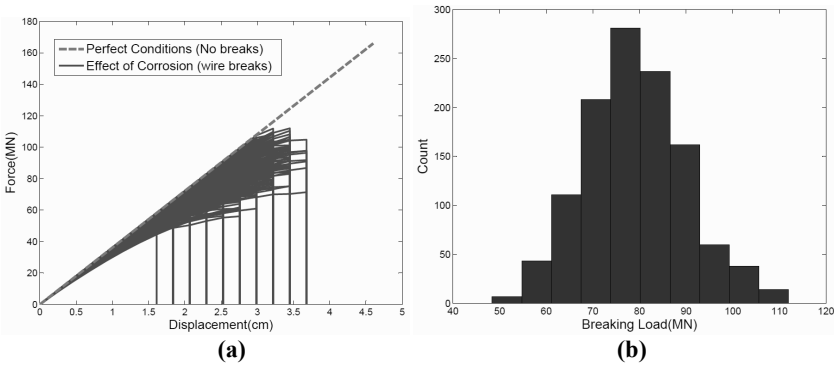
Figure 6. 1000 realizations: (a) Force-displacement curve and (b) histogram

REFERENCES

Boussinesq, J. (1885). *Appications des potentials a l'étude de l'équilibre et du mouvement des solides élastiques*.Gauthier-Villars.

Brigham, E.O. (1988). *The Fast Fourier Transform and its Applications*. Prentice Hall, New Jersey.

Gronquist Bridsall Steinman, Boynton and Columbia University, 1998. *Williamsburg Bridge cable investigation program: Final report*. Submitted to the New York State Dept of Transp. and New York City Dept. of Transportation, New York.

Mateo, J., Deodatis, G. and Billington, D. (1994) . "Safety analysis of suspension bridge cables:Williamsburg bridge. " *ASCE-Journal of Structural Engineering*, 120(11):3197-3211.

Montoya, A., Waisman, H. and Betti, R. (2012). " A simplified contact-friction methodology for modeling wire breaks in parallel wire strands." *Computers and Structures*, 100-101:39-53.

Shi, Y., Deodatis, G. and Betti, R. (2007). "Random field-based approach for strength evaluation of suspension bridge cables." *ASCE-Journal of Structural Engineering*, 133(12):1690-1699.

Shinozuka, M. and Deodatis, G. (1996). "Simulation of multi-dimensional gaussian stochastic fields by spectral representation." *Applied Mechanics Rev.*, 49(1):29-52.

Shinozuka, M. and Jan, C. M. (1972) . "Digital simulation of random processes and its applications." *J. Sound and Vibration*, 25, 111–128.

Suzumura, K. and Nakamura, S. (2004). "Environmental factors affecting corrosion of galvanized steel wires." *ASCE-Journal of Materials in Civil Engineering*, 16(1):1-7.

Waisman, H., Montoya, A., Betti, R. and Noyan, I.C. (2011). " Load transfer and recovery length in parallel wires of suspension bridge cables." *ASCE-Journal of Engineering Mechanics*, 137(227).

Wriggers, P. (2006). *Computational contact mechanics*. 2nd edition, Springer, Heidelberg.

Cyberinfrastructure Middleware and Analytical Tool Sets for Automated Mining of Massive Structural Health Monitoring Data Sets

Y. Zhang[1], S.M. O'Connor[1], J. P. Lynch[1,2], G. van der Linden[3] and A. Prakash[2]

[1]Dept. of Civil and Env. Engineering, Univ. of Michigan, Ann Arbor, MI 48109
[2]Dept. of Electrical and Computer Eng., Univ. of Michigan, Ann Arbor, MI 48109
[3]SC Solutions, Inc., 1261 Oakmead Parkway, Sunnyvale, CA, U.S.A 94085

ABSTRACT

Deployment of dense networks of low-power wireless sensors has been shown to be a cost effective approach to structural monitoring that can generate massive volumes of data. Building a specialized cyberinfrastructure system is an efficient way to store and organize large volumes of data (sensor and metadata) in addition to processing it. Availability of structure metadata (*e.g.*, geometric details, material properties, inspection histories) further enhances the post-collection analysis of the data collected. In this paper, a comprehensive cyberinfrastructure and its associated computing technologies are proposed to serve as the backbone of large-scale permanent structural monitoring systems. A powerful data server called *SenStore* is described as the primary building block of the proposed cyberinfrastructure framework. This paper describes *SenStore* and its use in continuously monitoring the Telegraph Road Bridge (Monroe, MI) using a permanently installed wireless sensor network.

INTRODUCTION

Dense low-power wireless sensor networks (WSN) have been proven effective over the past decade in offering the ability to serve as long-term structural monitoring systems for the evaluation of structural performance and health *(Lynch et al.* 2009). While such systems offer tremendous promise, they have the potential to produce incredibly large volumes of sensor data. Comparatively less work has been conducted on data management and processing architectures that can harness the full potential of sensor data collected. Furthermore, to extract meaningful information from the data, decision makers still urgently require computational tools that can automate the processing of data to make structural assessments for the management of risk. A secure cyberenvironment, customized for structural health monitoring applications, is required to manage not only large volumes of sensing data, but also structural metadata (*e.g.*, the structure's geometric details, material properties, inspection history). Previously, several tools have been developed by researchers that begin to address these needs including sensor data storage using single relational database systems (Koo *et al.* 2011; Cross *et al.* 2013; Allen *et al.* 2003; *Magalhães et*

al. 2012). In this study, a fresh new approach to the design of a cyberinfrastructure framework is proposed for wireless structural monitoring systems. The cyberinfrastructure architecture is introduced as a general framework for the scalable storage of large volumes of sensor data and structural metadata while offering secure application programming interfaces (APIs) that allow distributed data processing tools to gain access to the system data. The system is designed to be compatible with different sensor networks and structures. At the core of the cyberenvironment is *SenStore*, a server platform that includes a unique database design for scalable management of structural monitoring data. *SenStore*'s server-client model is applied to offer secure access to stored data via its rich API. Multiple automated data analysis tools have been developed for the *SenStore* platform to evaluate the sensing network's performance, provide web database viewing and to perform sophisticated structural analyses (Kurata *et al.* 2012; Zhang *et al.* 2012). This paper presents the cyberenvironment and its associated details while illustrating its application to bridge monitoring applications using the Telegraph Road Bridge (TRB).

CYBERINFRASTRUCTURE DESIGN FOR STRUCTURAL MORNITORING

In a long-term structural monitoring system, continuous post-collection data processing of large data sets is critical to monitoring the performance of the instrumented structure in real-time or near real-time. Therefore, cyberinfrastructure middleware is introduced to provide a well-organized computing environment to improve the efficiency of the post-processing of sensor data in an autonomous manner. At the core of the cyberinfrastructure system's design is a powerful server architecture termed *SenStore*. *SenStore* is a server platform which includes a database and data management system designed explicitly for structural monitoring applications. The database consists of two linked database systems within the same architecture (as shown in Figure 1): PostgreSQL and HDF5. PostgreSQL is selected to store structural metadata because it is an objective oriented, open source relational database that supports rich data types. In the latest version of *SenStore*, the PostgreSQL schema includes information such as structural details (*e.g.*, defines components, geometry, materials), finite element model information (*e.g.,* coordinates, elements), structural monitoring system deployment details (*e.g.,* sensor locations, sensor types), sensing events, and inspector assessment reports. However, the relational database is not well suited for storing large volumes of time history sensor
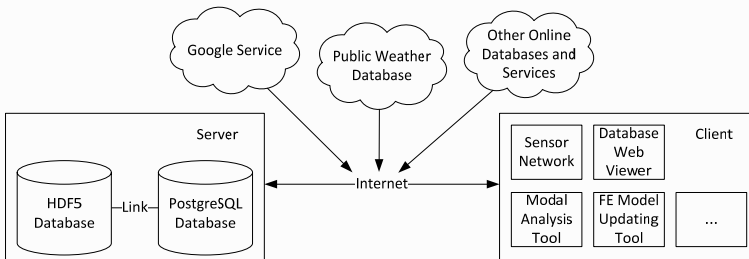
Figure 1: Architectural overview of the *SenStore* cyberenvironment

data and could represent a performance bottleneck during routine queries. To achieve efficient input/output data flow, *SenStore* integrates HDF5, an efficient data library system optimized for time-history data, to store large volumes of sensor data with time stamps (*e.g.*, raw data collected from structural monitoring sensors). HDF5 features fast read/write functionality for massive data repositories, especially when accessing random portions of data. Connections are implemented from within the PostgreSQL database to link explicitly to the sensor data housed in HDF5. This strategy allows all user queries to be handled by PostgreSQL. The consolidation of structural metadata and sensing data into one centralized database system transforms *SenStore* into a scalable platform upon which autonomous data analysis can be performed autonomously in near real-time.

In a post-9/11 world, structural metadata and sensing data are highly sensitive; hence, data security becomes a dominant issue. *SenStore* provides secure access to the server from remote clients via a secure server-client model. *SenStore* takes advantage of ZeroICE (www.zeroc.com) server technology to offer clients access to bridge data through an indirect connection to the *SenStore* database using prepared statements in the client application programming interface (API). A server manager is built on the server to arrange for the function calls from the clients with detailed client request information logged for future reference. The API also allows users to query the database with prepared query statements thus preventing random queries that may have a malicious intent associated with them. Finally, the cyberinfrastructure framework proposed also offers users opportunity to profit from third-party web services (*e.g.,* Google Cloud) and public databases (*e.g.,* weather or traffic information databases).

LONG-TERM WIRELESS BRIDGE MONITORING SYSTEM

The Telegraph Road Bridge (TRB) is a multi-girder steel concrete composite bridge located in Monroe, MI. The bridge is comprised of three main spans and seven beam lines and has a skew angle of 57 degrees. The end spans are fixed at the abutment end and roller supported with rocker bearings atop interior piers. The main span is completely suspended to the end spans by link plate assemblies at each beam line, also referred to as pin-and-hanger connections. The TRB monitoring system is a two tier entity. The lower tier consists of a wireless sensor network utilizing the *Narada* wireless sensing platform. The upper tier includes the *SenStore* database management system (located in Ann Arbor, MI). Data is first stored locally at the TRB by the wireless monitoring system using a Linux server implemented on a low-power PC-104 single board computer. A wireless transceiver is connected at the server to provide two-way communication with the wireless sensors. Stored data is relayed via a 3G cellular network for storage in the *SenStore* system.

Several deterioration processes were of interest in assessing the performance of the TRB. A sensor network using 31 *Narada* nodes was installed to collect response data from 15 accelerometers, 36 strain gages, and 6 thermistors deployed throughout the TRB (shown in Figure 2). Each *Narada* node is also sensing the voltage level of its rechargeable battery in order to ensure each node has sufficient power. This sensing strategy is designed to: 1) facilitate modal analyses for finite
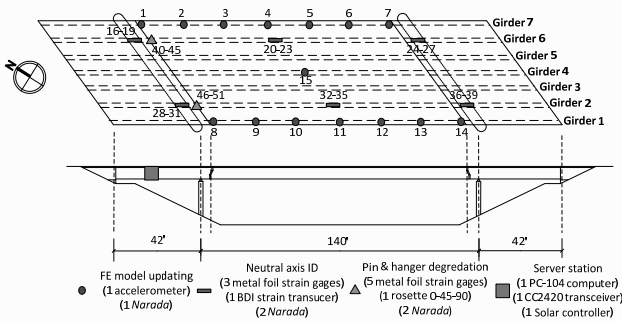
Figure 2: Sensor deployment on the Telegraph Road Bridge

element model updating; 2) identify neutral axis position of the girder-deck superstructure for composite action assessment; and 3) monitor link plate strains to identify corrosion in the pin-and-hanger connections as well as perform fatigue analyses on the link plate and pin. The permanent deployment of the TRB monitoring system is enabled by powering the system solely through solar energy. A more detailed deployment plan can be referred in O'Connor, *et al.* (2012). The system is configured to perform data acquisition at 200Hz for 1 minute every 4 hours. As there are 88 sensing channels installed, nearly 38 MB sensing data is generated every day (14 GB a year).

ANALYTICAL TOOLSET FOR PROCESSING SENSOR DATA

Multiple analytical tools have been developed to assist in the visualization, administration and processing of the sensor data (Zhang, *et al.* 2012). These tools are contained in data processing servers that remotely access sensor data and bridge metadata warehoused in the *SenStore* server. In this study, two major data processing tasks are illustrated using *SenStore* simply to highlight its impressive capabilities: 1) automated modal characteristic extraction from sensor data and 2) model updating of finite element models automatically generated from bridge metadata.

Multivariable Modal Analysis Tool

Modal properties (*i.e.*, modal frequency, damping ratio and mode shapes) are an important global property of a structure. These properties are often used for model updating of finite element analysis methods. Tools have been developed to compute modal parameters autonomously using bridge acceleration data. Frequency domain decomposition (FDD) (Brincker *et al.* 2001) and stochastic subspace identification (SSI) (Peeters and De Roeck 1999) are both adopted in this study. After querying data from *SenStore*, modal frequencies and corresponding mode shapes are extracted from
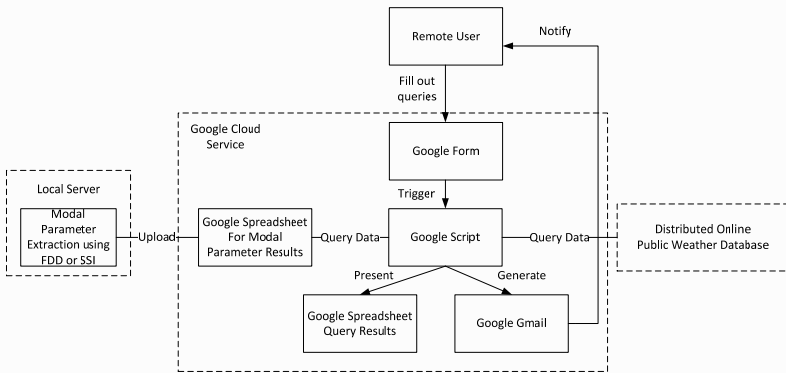
Figure 3. Cloud-supported multivariable modal analysis tool

clients running robust implementations of the FDD and SSI methods; damping ratio can only be obtained from the SSI method.

The modal frequencies are known to vary (as much as 5%) as a function of changes in the environment (*e.g.,* temperature, wind speed) and traffic. In fact, these environmental influences are often more significant than the influence caused by structural damage on a large-scale structure (Brownjohn, *et al.* 2010; Hua, *et al.* 2007). To highlight the capability of *SenStore* to evaluate large volumes of bridge response data, this study employs FDD and SSI data processing clients to extract modal frequencies to correlate them against environmental data such as bridge temperature. The client created to perform this task is supported by the Google Cloud Platform with several Google services applied to perform the computation on the cloud. Figure 3 shows the conceptual flow chart of the proposed cloud-supported multivariable modal analysis tool once the modal information has been extracted from *SenStore*'s database by the FDD or SSI client. A set of Google scripts are triggered by users' requests which are entered on a Google Form; the users specify the time period to study, the structure (*e.g.,* Telegraph Road Bridge), the modal analysis method (*e.g.*, FDD or SSI) and environmental parameters to consider (*e.g.*, temperature). The form also requests the users e-mail for notification. Modal frequency results stored on a Google Spreadsheet are selected for each measurement cycle and computation is performed to obtain the corresponding environment parameter using APIs provided by public weather databases (*e.g.* http://www.wunderground.com/). The results, including a plot of the specified modal frequencies versus time or desired environment parameter are created along with a table of numerical values. These results are then presented to the user in the format of Google Spreadsheet and Google Chart. The results from this multivariable modal analysis tool using the FDD method on two months (June and July 2012) of TRB data are shown in Figure 4 (which is the output from Google Charts). The modal frequencies of TRB vary about ±3% near a mean value (2.455 Hz for Mode 1). However, when investigating the relationship between temperature and modal
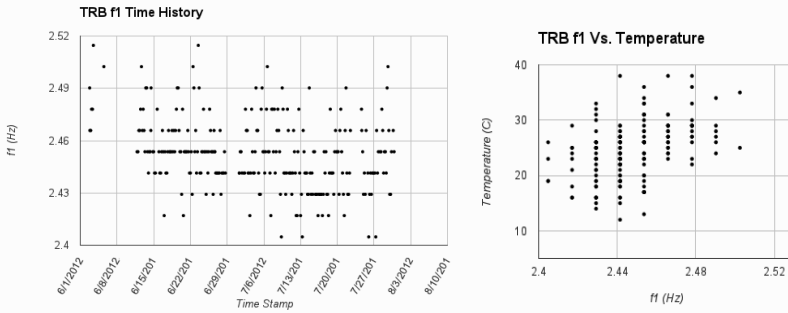
Figure 4: Results from the automated multivariable modal analysis: TRB modal frequency time history (left) and as a function of temperature (right).

frequencies, it is very clear the modal frequency of TRB has a fairly linear relationship with temperature.

Automated Finite Element Analysis Driver/Interface

Finite element analysis (FEA) is an important structural analysis method in both the design and management of complex structures. Most commercial large-scale FEA is performed using commercial software (*e.g.,* SAP, ABAQUS, ADINA) via graphic user interfaces that facilitate manual data entry. Taking advantage of the marriage of sensor data generated from the long-term structural monitoring system and metadata containing information about the structure, automated FEA model updating can be performed with the assistance of an FEA model updating client invoking a commercial FEA platform. Since *SenStore* stores rich metadata about the monitored structure in a manner specific to the finite element method, a FEA model can be automatically extracted from the *SenStore* PostgreSQL database by mapping the database schema to an input file format used by a commercial FEA platform (in this study, SAP).   A client has been authored to make this mapping using user-defined control parameters (*e.g.,* output results type) with the overview of the client operation shown in Figure 5.   Figure 6 shows a small portion of the input file (generated by the FEA client) and output results (from SAP) when a user specifies "*Modal Periods And Frequencies*" analysis.   This tool is designed in an object oriented way that it could be easily adapted to different bridges and to different FEA software. It is written in Python and is designed to extract data from the *SenStore* database, create a SAP input file and to run the FEA software using the input file.

CONCLUSION

SenStore takes advantage of the common characteristics between different structural monitoring systems and provides an efficient and well-organized cyberenvironment for the development of automated data processing tools. The specialized relational *SenStore* database creates a complete data schema for structural
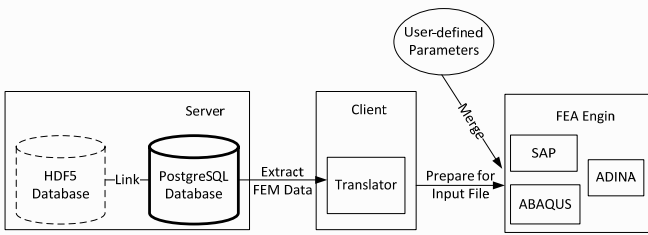
Figure 5: Flow chart of automated finite element client.

(a)

TABLE: Modal Periods And Frequencies						
OutputCase	StepType	StepNum	Period	Frequency	CircFreq	Eigenvalue
Text	Text	Unitless	Sec	Cyc/sec	rad/sec	rad2/sec2
Modal	Mode	1.0000	0.4384	2.2812	14.3333	205.4427
Modal	Mode	2.0000	0.4136	2.4177	15.1909	230.7630
Modal	Mode	3.0000	0.3766	2.6557	16.6861	278.4262
Modal	Mode	4.0000	0.2193	4.5597	28.6492	820.7744
Modal	Mode	5.0000	0.1287	7.7681	48.8084	2382.2641

(b)

Figure 6: Screenshot of the client generated input file for SAP and corresponding output file containing modal analysis results for the TRB

monitoring systems for bridges and the *SenStore* server-client model improves the security of the data. The current suite of autonomous distributed tools serves different purposes and could be easily adapted from one monitoring system to the other. The multivariable modal analysis tool assists researchers in developing the relationship between modal parameters and environmental factors using the Google Cloud platform. Moreover, the automated finite element analysis interface largely increases the efficiency to perform finite element analysis by computing the relationship of database schema and FEM mechanism.

ACKNOWLEDGMENTS

REFERENCES

Allen, D. W., Clough, J. A., Sohn, H., and Farrar, C. R. (2003). "A Software Tool for Graphically Assembling Damage Identification Algorithms." *SPIE 5057, Smart Structures and Materials: Smart Systems and Nondestructive Evaluation for Civil Infrastructures*, San Diego, CA.

Brincker, R., Zhang, L., and Andersen, P. (2001). "Modal identification of output-only systems using frequency domain decomposition." *Smart Materials and Structures*, 10(3), 441–445.

Brownjohn, J.M.W., Magalhaes, F., Caetano, Elsa, and Cunha, A. (2010). "Ambient vibration re-testing and operational modal analysis of the Humber Bridge." *Engineering Structures*, Elsevier Ltd, 32(8), 2003–2018.

Connor, S. M. O., Lynch, Jerome P, Ettouney, M., Linden, Gwen Van Der, and Alampalli, S. (2012). "Cyber-enabled Decision Making System for Bridge Management using Wireless Monitoring Systems : Telegraph Road Bridge Demonstration Project." *ASNT - NDE/NDT for Highways and Bridges*, New York, NY.

Cross, E. J., Koo, K.Y., Brownjohn, J.M.W., and Worden, K. (2013). "Long-term monitoring and data analysis of the Tamar Bridge." *Mech. Systems and Signal Processing*, Elsevier, 35(1-2), 16–34.

Hua, X. G., Ni, Y. Q., Ko, J. M., Asce, F., and Wong, K. Y. (2007). "Modeling of Temperature – Frequency Correlation Using Combined Principal Component Analysis and Support." *Journal of Computing in Civil Engineering*, 21(April), 122–135.

Koo, Ki Young, Battista, N. De, and Brownjohn, James M W. (2011). "SHM Data Management System Using MySQL Database with Matlab and Web Interfaces." *Structural Health Monitoring of Intelligent Infrastructure*, Cancun, Mexico.

Kurata, M., Kim, J., Lynch, J. P., Linden, G. W. Van Der, Sedarat, H., Thometz, E., Hipley, P., and Sheng, L. (2012). "Internet-Enabled Wireless Structural Monitoring Systems: Development and Permanent Deployment at the New Carquinez Suspension Bridge." *Journal of Structural Engineering*, (March), 462.

Lynch, Jerome P., Kamat, V., Li, V. C., Flynn, M., Sylvester, D., Najafi, K., Gordon, T., Lepech, M., Emami-Naeini, A., Krimotat, A., Ettouney, M., Alampalli, S., and Ozdemir, T. (2009). "Overview of a cyber-enabled wireless monitoring system for the protection and management of critical infrastructure systems." (H. F. Wu, A. A. Diaz, P. J. Shull, and D. W. Vogel, eds.), 7294(734), 72940L–72940L–11.

Magalhães, F., Amador, S., Cunha, Á., and Caetano, E. (2012). "Dynamo – software for vibration based Structural Health Monitoring." *International Conference on Bridge Maintenance, Safety and ManagementBridge Maintenance*, Stresa, Lak Maggiore, Italy.

Peeters, B., and De Roeck, G. (1999). "Reference-based Stochastic Subspace Indentification for Output-only Modal Analysis." *Mechanical Systems and Signal Processing*, 13(6), 855–878.

Zhang, Y., Kurata, Masahiro, Lynch, Jerome P., Van der Linden, Gwendolyn, Sederat, H., and Prakash, A. (2012). "Distributed Cyberinfrastructure Tools for Automated Data Processing of Structural Monitoring Data." A. L. Gyekenyesi, ed., San Diego, CA, 83471Y–83471Y–8.

Spatiotemporal Dimensions of Network Density-based Clustering for Water Pipe Maintenance

L. T. Ong[1], J. M. VanBriesen[1] and J. H. Garrett, Jr.[1]

[1]Department of Civil and Environmental Engineering, Carnegie Mellon University, 119 Porter Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213; PH (412) 268-2940

ABSTRACT

In the U.S., many drinking water assets operate beyond their designed lifetimes, and constrained resources necessitate optimizing repair and replacement. For pipe infrastructure, clustering tools can characterize their susceptibility to failure by identifying relationships among descriptive or measured features. In these complex systems, algorithmic learning approaches can provide a first insight before expert knowledge is applied, reducing time and labor. The state-of-the-art techniques often rely upon static characteristics. In this work, pipe maintenance records are analyzed through the network OPTICS ("Ordering Points To Identify the Clustering Structure"), which forms a hierarchal density-based clustering structure. The study compares and extends OPTICS to a temporal context; exploring the evolution of clustering structure provides additional insight. The findings suggest this spatiotemporal approach is applicable for improved asset management.

INTRODUCTION

Ensuring the quality of drinking water requires proper maintenance and improvements of pipe networks. In the U.S., many assets are beyond their designed lifetimes and constrained resources necessitate optimizing asset management (US EPA, 2002). Water distribution systems require large investments, and, when breaks and leaks occur, have costs beyond only repair or replacement. There is major value in prioritizing upgrades and inspections to prevent failure events.

Decision support systems (DSS) have been proposed to help utilities manage assets. Deterioration models may assist in predicting break rates, but such models cannot predict failures for specific pipes or account for changes to features (Wang et al., 2009). A spatial tool using a Geographical Information System (GIS) employs fuzzy inference to examine features in an area, but requires calibration to set multiple attribute thresholds and weights (Sinske and Zietsman, 2004). More overarching, an agent-based framework proposes multiple agents to exchange and act on information according to specific roles, such as data cleaning or prediction (Davis, 2000). DSS are meant to be interactive and promote efficient decision-making. Yet, they are not meant to substitute for expert knowledge (Saldarriaga et al., 2009).

A significant challenge is using large data sets that come from many sources, which are often stored in legacy information systems. Algorithmic data-mining

techniques are well-suited to the task, and this paper examines the viability of a clustering approach for this domain. As an unsupervised learning method, it assists in creating a structure to explain features inherent in the data. The goal is to assign observations to a set of groups based on similar characteristics or divide observations based on dissimilarities. With regard to one data-mining methodology, this work addresses the stages of data exploration, pattern generation, and enables pattern monitoring (Davis, 2000). These steps provide initial insight before expert knowledge is brought in to deepen the analysis.

RELATED WORK

For discovering patterns, unsupervised approaches preprocess data from unlabeled training data into groups based on dis/similarity, which can be subsequently analyzed. Clustering algorithms used for classification face several challenges: minimizing domain knowledge needed for input parameters, finding clusters of arbitrary shape, and computational efficiency for large databases (Ester et al., 1996). Density-based clustering addresses these issues by recognizing areas of higher point density relative to the surrounding region to identify clusters.

The popular DBSCAN (density-based spatial clustering of applications with noise) is a hierarchal clustering method that defines clusters using *density connectedness* (where two points are connected if within a certain distance and density threshold) and *density reachability* (where two points are reachable if through point(s) that are density-connected) (Ester et al., 1996). Points that are not density-reachable are considered noise or border points, which give DBSCAN and its variations the benefit of outlier detection (Hinneburg and Keim, 1998). OPTICS (ordering points to identify the clustering structure) is a modification of DBSCAN to assist in analyzing clusters in data with varying density (Ankerst et al., 1999). It overcomes the single-link effect, where clusters are not separated because they are linked by a few points having small differences in distance, and the challenge of comparing clustering from varying density parameter settings. It orders points so that points that are spatially closest are neighbors, i.e. clusters can be visualized in a plot of the reachability distances of each cluster's members. Points within the cluster are again separated by differences in density, which can be seen in the reachability plot.

The minimum cluster size, *MinPts*, and neighborhood distance, ε, are the only two parameters that need to be specified. A detailed explanation of OPTICS will not be covered here, and the reader is referred to Ankerst et al. 1999 for a complete treatment. It has been adapted to use network path as a distance metric in order to create clusters that reflect physical geometries (de Oliveira et al., 2011). The method uses Dijkstra's shortest path algorithm in order to obtain the distances of paths in the ε-neighborhood around point events, instead of Euclidean distances.

METHODOLOGY

The metropolitan water distribution system used as the case study was considered to be a network with no loops and un-directed flow. In reality, the physical layout, pressure and gravity likely create directed flows, but a calibrated hydraulic model, which would include this information, was not available. While a flow-

directed model would provide additional input, assuming an undirected network preserves more of the possible paths between points in the system, enabling generalization. For this network, it is unusual for a node to be connected to 4 or more other nodes (imagine the physical difficulty of connecting many water pipes together at a single junction); over 75% are connected to only one other pipe. Network distance clusters were thus expected to be spread out and irregular in shape according to pipe geometries.

For pipe break and leak events, network OPTICS clustering and Euclidean hierarchal clustering were compared on a section of the system. Approximately 1,400 events from 2005-2008 were clustered using network distances as described previously. The Euclidean version creates a multilevel binary hierarchy such that every event belongs to some cluster that is eventually linked together at some level. The algorithm works by 1) calculating the Euclidean distance between each pair of points, 2) grouping objects based on a linkage function, or method of computing distance between clusters, and 3) pruning the tree into clusters. The linkage function was chosen based on the highest *cophenetic correlation coefficient*, which was the unweighted average distance (c = 0.8282). It is a measure of how well the hierarchy fits the dissimilarity, or distance, between events. Pruning was accomplished by specifying the number of clusters desired.

On a subsection of the network, both types of clustering were performed. Using a minimum cluster size threshold of 5 point events, network OPTICS produced 48 non-noise clusters, with 42 being leaf clusters, or not containing sub-clusters. Hierarchal clustering in Euclidean space agglomerates data into clusters depending on either a threshold of the distance between clusters or the maximum number of clusters. The latter criterion, $maxClusters_E$, was selected and varied since there is no previous knowledge of appropriate parameters for this system. As $maxClusters_E$ increases, the wider clusters fragment regardless of membership size (Figure 1).
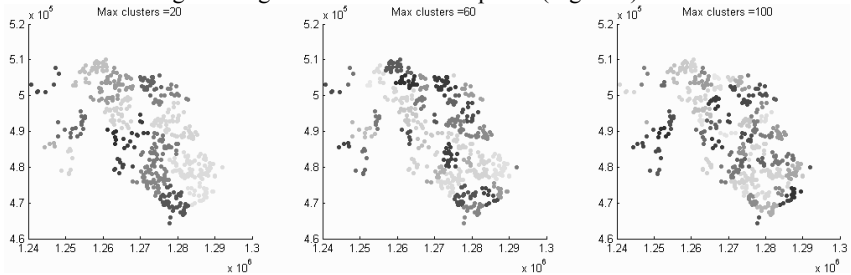
Figure 1. Hierarchal Euclidean clustering of pipe events.

Hierarchal clustering using Euclidean distances highlighted areas that were not considered clusters using network OPTICS. Since this type of clustering does not distinguish noise, many clusters found using hierarchal clustering are likely not significant. With four years of data and parameter $maxClusters_E = 238$, over half of the clusters include less than the minimum size considered for OPTICS. Sensitivity of the parameter is shown in Figure 4. To compare with OPTICS, $maxClusters_E = 47$ such that also 42 clusters are formed containing 5 or more pipe events.

The minimum cluster size for network OPTICS was selected to be *minPts* = 5, though varying this parameter incrementally did not appear to change the results much. Neighborhood distance was set at ε = 11,000 meters, which is approximately three times greater than the average of the lengths of the top ten longest pipes. A longer ε distance creates more clusters that extend across the network, whereas a smaller ε produces fewer clusters. ε should at least be longer than the longest lengths, or else some pipes may never be traversed by the algorithm, effectively eliminating those network connections.

Afterwards, an extension to temporal dimensions was examined in two ways. The first method was across specific months, and the second according to a moving time window. In this case, Oct-Mar and Apr-Sep were used to arbitrarily capture climate, but may be divided more granularly based on real weather data. In some studies, frost, temperature, or moisture levels were significant influences (Kleiner and Rajani, 2002). To analyze time-space aspects, a window of 500 events was incremented in steps of 200 events to demonstrate how clusters vary.

RESULTS AND DISCUSSION

Non-spatial analysis. Characterizing event types based on only non-spatial features serves as a baseline to reference with the results of cluster-based analysis. For the pipe system, the available features are pressure zone, diameter and material. Cast iron pipes (which have been discontinued as a choice since 1977) are known for this network to be responsible for 98% of water main breaks. Knowing this, material type won't be considered as a distinguishing factor. Examining failure types, pipe main breaks occur most in 6" (37%), 8" (35%), and 3" (12%) diameter pipes. Leaks occur more frequently in 6" (22%), 8" (20%), and 3" (15%) diameter pipes (but 17% are indeterminate). Though 3" pipes make up less than 5% of the pipes, the frequency of events on pipes that size are greater than pipes of larger, more common sizes.

The test section of the network consists of 1,398 events that occurred in the six months of from 2005 to 2008, over 171,000 pipes. The population of pipe sizes is similar in proportion to the whole network, whereas the material composition is different. The top four sizes and materials are listed in Table 1 as a percentage of the total test network. Also, the pipe events are broken down into the most common characteristics, although 10.4% of pipe events occurred in pipes with 3" diameters, more than in 12" and 10" pipes as similar to the entire network.

Table 1. Proportion of pipes and pipe events by feature.

Diameter	% of Total	% of Events	Material	% of Total	% of Events
8"	35.0%	32.3%	Cast Iron	65.2%	84.0%
6"	24.6%	34.6%	Ductile Iron	31.7%	15.0%
12"	9.8%	6.1%	Concrete	2.7%	0.8%
10"	8.5%	6.4%	Steel	0.2%	0.1%

Spatial analysis. Of the significant clusters, the methods create distinct clusters of size and shape. In region A, OPTICS creates one cluster whereas Euclidean splits it into multiple clusters. Region B shows many more significant

clusters under Euclidean clustering. The clusters in Region C differ significantly as well, with many of the Euclidean clusters considered noise under the network scheme.
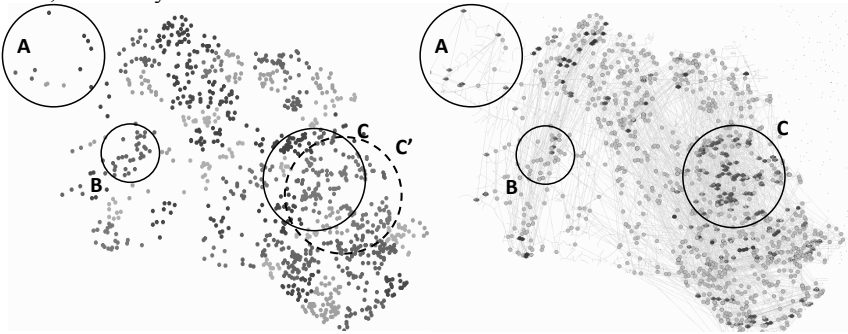
Figure 2. Some regional differences between Euclidean clustering (Left) and Network OPTICS (Right) for pipe events, 2005-2008. Grey region in OPTICS map denotes the noise cluster.

The largest network OPTICS cluster contains 51 pipe events (center of Region C on right panel of Figure 2). The affected pipes are all in the same pressure zone, made mostly of material type "I" (84%) with a diameter of 6" (37.3%) or 8" (39.2%). The comparable cluster C in the other scheme (left panel of Figure 2) encompasses many of the same events, which is expected given the almost grid-like density of pipes, and hence, network paths not being that different from line-of-sight distances. However, the Euclidean cluster in Region C' additionally incorporates events that do not share many connections, as the centroid is located in a space with no pipes.

Comparing the clusters created from specific months, the cluster in Region C is the only cluster that persists according to OPTICS. Many of the smaller clusters found in the colder months are not significant in the warmer months (Figure 3).

Figure 3. Clusters found using network OPTICS for a) Oct-Mar, and b) Apr-Sep.

Generally, the number of significant clusters is reduced for network OPTICS, as expected by the smaller population of events. Considering Oct-Mar (885 events), 25 non-noise clusters (20 leaf clusters) are found. Considering the other half (513 events), 10 non-noise clusters (8 leaf clusters) are found. Selecting specific timeframes impacts Euclidean clustering by increasing its sensitivity to the

maxClusters$_E$ parameter (Figure 4). Depending on the sample size, there is an optimal region of maximizing the number of significantly sized clusters.
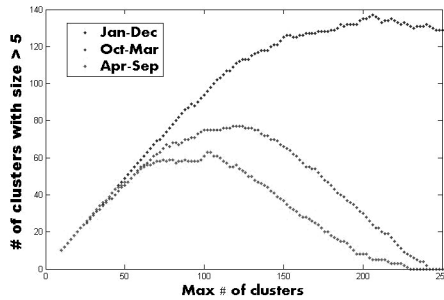
Figure 4. Sensitivity of cluster size to max number of clusters, *maxClusters$_E$* for a) Annual data, b) Oct-Mar, and c) Apr-Sept.

Focusing on Region C, OPTICS produces several leaf clusters, as denoted by separate colors. The reachability plot illustrates the network distance between neighboring pipe events; short between-event distances are clustered, with divisions created where there is a significantly longer between-event distance (Figure 5). In many cases, the distance is considered unreachable by the parameter threshold (black areas). In comparison with the Euclidean method, the central locations of many clusters in this region are the same, but membership and overall shape are non-compact and generally smaller due to many events deemed as noise.
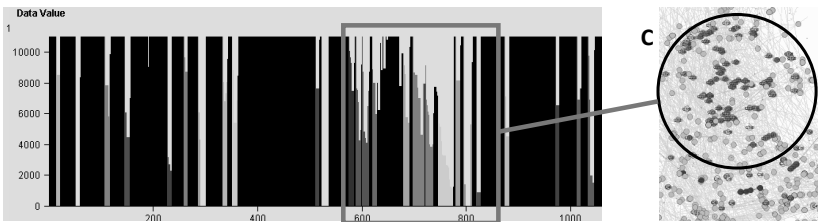
Figure 5. Events plotted by density distance highlighting Region C, 2005-2008.

Over a moving time horizon, the clusters change in size. They may also separate and combine. Using the reachability plots for selected time windows, the cluster(s) identified in Region C can be shown to evolve (**Figure 6**). In some intervals, a cluster appears to merge in t = 2 to t = 3 and t = 4 to t = 5 and separate in t = 3 to t = 4. Many of the smaller clusters pop up but tend to disappear, though in t = 4 there are regions in the noise (black) cluster that have a density that is not infinity, which appear to evolve into clusters once events have occurred in that area in the next increment. In these examples, the test network demonstrates that network OPTICS can potentially track the activity of clusters as they evolve and as events become deemed part of a significant, non-noise-associated cluster.
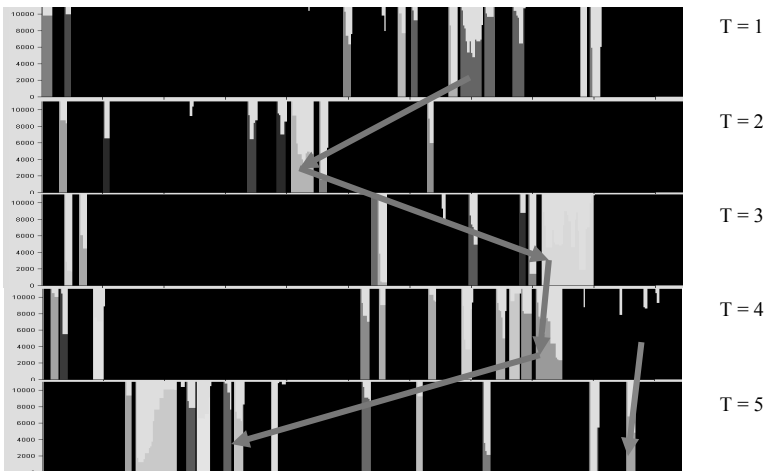
Figure 6. Reachability plots showing cluster evolution over T steps.

CONCLUSION

Network OPTICS presents a promising method of extracting clustering behavior from complex pipe systems. It demonstrates different cluster shapes especially in areas of sparse connectivity. Also, cluster sizes were significantly smaller; distinguishing noisy events greatly reduced cluster sizes over a range of parameters. The advantage that network OPTICS has over Euclidean clustering is the ability to extract non-compact cluster shapes where a Euclidean cluster would not reflect the physical geography of the system. Moreover, OPTICS can show how clusters change by growing, shrinking, merging and separating, which is a key challenge in cluster tracing (Günnemann et al., 2011). However, without more detailed features, the reasons for cluster behavior were not explained for this system.

Still, there is a significant challenge with determining parameters. When considering subsets of pipe events, large adjustments are needed for both methods. Choosing parameters that are too large reduces the number of clusters identified. For practicality, clustering using small minimums would result in a number of clusters that would be time-consuming to examine for potential causes of a break or leak.

Further work on defining evolving behavior would be valuable for increasing the usefulness of clustering point events. In contrast to strictly examining static features, it presents an opportunity to incorporate time-based feature data in a spatial framework. For maintaining pipe networks, this would be valuable in order to incorporate water quality or pipe condition measurements, for instance. The challenge of clustering efficiently and resilient to noise or bad data requires future research, as well as understanding the parameters to adapt the method to different applications.

ACKNOWLEDGMENTS

REFERENCES

Ankerst, M., Breunig, M.M., Kriegel, H., and Sander, J. (1999). OPTICS: Ordering Points To Identify the Clustering Structure. In Proceedings of ACM SIGMOD International Conference on Management of Data, (Philadelphia, PA), pp. 49–60.

Davis, D.N. (2000). Agent-based decision-support framework for water supply infrastructure rehabilitation and development. Computers, Environment and Urban Systems *24*, 173–190.

Ester, M., Kriegel, H.P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, pp. 226–231.

Günnemann, S., Kremer, H., Laufkötter, C., and Seidl, T. (2011). Tracing evolving clusters by subspace and value similarity. Advances in Knowledge Discovery and Data Mining 444–456.

Hinneburg, A., and Keim, D.A. (1998). An efficient approach to clustering in large multimedia databases with noise (Bibliothek der Universität Konstanz).

Kleiner, Y., and Rajani, B. (2002). Forecasting Variations and Trends in Water-Main Breaks. Journal of Infrastructure Systems *8*, 122–131.

de Oliveira, D.P., Neill, D.B., Garrett, J.H., and Soibelman, L. (2011). Detection of Patterns in Water Distribution Pipe Breakage Using Spatial Scan Statistics for Point Events in a Physical Network. Journal of Computing in Civil Engineering *25*, 21–30.

Saldarriaga, J.G., Ochoa, S., Cortés, O.J., and Niño, C.A. (2009). Water Distribution Systems Corrective Maintenance Supported by Real-Time Use of Hydraulic Models. In World Environmental and Water Resources Congress 2009, pp. 1–11.

Sinske, S., and Zietsman, H. (2004). A spatial decision support system for pipe-break susceptibility analysis of municipal water distribution systems. Water SA *30*, 71–80.

U.S. Environmental Protection Agency (2002). The Clean Water and Drinking Water Infrastructure Gap Analysis.

Wang, Y., Zayed, T., and Moselhi, O. (2009). Prediction models for annual break rates of water mains. Journal of Performance of Constructed Facilities *23*, 47–54.

Interpreting the Dynamics of Embankment Dams through a Time-Series Analysis of Piezometer Data Using a Non-Parametric Spectral Estimation Method

In-Soo Jung[1], Mario Berges[1], James H. Garrett Jr[1], Christopher J. Kelly[2]

[1]Carnegie Mellon University, Department of Civil and Environmental Engineering; 5000 Forbes Avenue, Pittsburgh, PA 15213; e-mail: {ijung, mberges, garrett}@andrew.cmu.edu
[2]US Army Corps of Engineers, Risk Management Center; 1000 Liberty Avenue, Suite 1107, Pittsburgh, PA 15222; email: christopher.j.kelly2@usace.army.mil

ABSTRACT

A common approach used by engineers to monitor and analyze data collected from piezometers installed in embankment dams is to generate time history plots and visually identify any spikes or anomalies in them. However, such practice has several limitations when capturing complicated relationships among a number of factors that affect piezometric readings. This is especially true when periodic or dominant variations that exist in time-series data are of concern, given that environmental and process noise can sometimes mask these variations. In this paper, we propose applying Moving Principal Component Analysis (MPCA) and Robust Regression Analysis (RRA), which have been shown to be successful in other applications, to extract relevant components and detect anomalies in piezometer measurements, which are one of the most important data to be monitored when evaluating the performance of embankment dams. The proposed anomaly detection method provides a more efficient way of understanding and detecting changes in piezometer data.

INTRODUCTION

There are more than 85,000 dams in the U.S., the majority of which were built decades ago. It is not surprising then that the number of deficient dams, as qualified by different evaluation methods, has increased in recent years. For example, according to the 2009 ASCE report card for the United States (U.S.) infrastructure, dams received an average grade of D, the lowest grade on that scale (ASCE 2009). Thus, more thorough inspections and immediate efforts are required to assess the condition of the dams and avoid any catastrophic consequences due to dam failures. In this study, we have focused specifically on one type of dams, namely embankment dams. These dams are the most common type of dam in use today (ASDSO and FEMA 2012). The most common aging scenario for embankment dams is internal erosion, which is mainly caused by seepage, and it is usually detected by periodic visual inspection and seepage measurement (USSD 2010). However, since it develops from the inside of dams, it is hard to be detected until it is too late.

When engineers monitor the performance of dams, they review instrumentation data and use additional sources, e.g., past inspection reports, construction photos, historical events, etc., as references. Especially when engineers are concerned about possible seepage problems, piezometer readings are closely analyzed. Piezometers measure the static water pressure at different points along the embankment, and these measurements are generally collected manually on a monthly or daily frequency, or automatically every few hours. When analyzing such data, time-history plots and/or correlation plots are often generated. The relationship between hydrostatic level and reservoir level can be understood using a piezometer hysteresis plot (ASCE 2000). If the piezometer level is directly influenced by the reservoir level with no other significant stimulus, the correlation plot will show a straight line (Gall 2007). Based on where the piezometers are located, i.e., the soil layers, distance between the piezometer and the reservoir, etc., the slopes of the correlation plots will vary due to different response times.

An example correlation plot is shown in Figure 1. The projection line is generated based on historic high pools, in which their corresponding maximum piezometer readings are included so that they can be compared with the previous piezometer responses. The projections are used as monitoring/action limits during future events to quickly verify that the piezometers are reacting in a predictable manner or whether additional monitoring is required or not. As the piezometer is influenced by other factors and lags behind the reservoir level, the data points will scatter along a sloped line, forming an elliptical envelope.
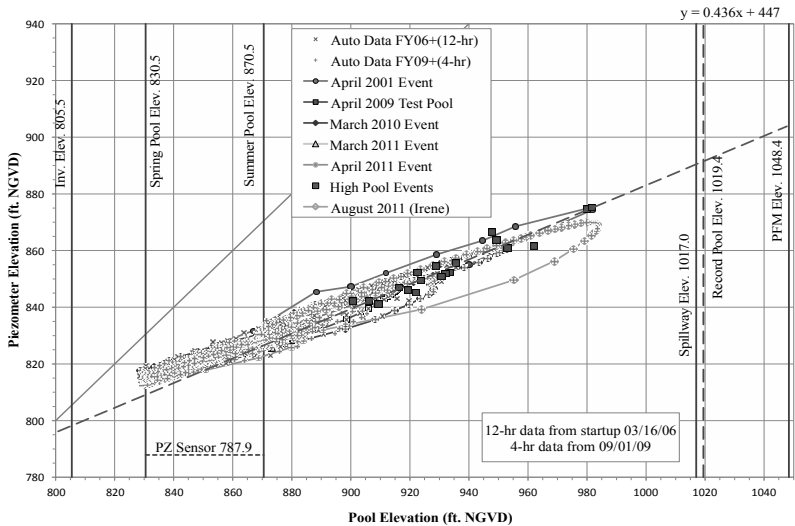
Figure 1. Correlation plot between piezometer and pool elevations

The major problem of analyzing such plots is that it is difficult to set the threshold values, or levels that need to be set around the straight line (or the envelope) to distinguish unexpected readings, or anomalies from normal readings.

Spikes in the piezometer data are rare, but since they are theoretically indicative of filter failure, piping (internal erosion), hydraulic fracturing, or other phenomena, it is difficult for engineers to dismiss them as bad data (Crum 2011). Thus, instead of visually identifying the data that deviate from the established norm, there should be a more quantitative and robust approach of detecting anomalies, not only to ensure dam safety, but also to reduce any subjectivity and efforts required by engineers.

There have been a number of data-driven anomaly detection methods considered in structural health monitoring (SHM) (e.g., Betti et al., 2006; Worden and Manson, 2007; Ying, 2012). Given the high dimensionality of many SHM datasets, as well as the complex relationships between the variables in the dataset, principal component analysis (PCA) has been widely applied and proved as a promising data analysis tool (e.g., Yan and Golinval, 2006; Yu et al. 2010; Tibaduiza et al., 2012). The main idea of PCA is computing eigenvectors associated with high variability of the data. Since the first few eigenvectors represent the directions of maximum variance or the variance of each independent component, the most dominant variation patterns can be captured. Due to this feature, observing changes in eigenvector structures over time can be adapted as an anomaly detection technique. For example, changes in these eigenvectors can be observed to detect if there have been any structural problems over time.

Many authors have applied this PCA-based method in combination with other statistical approaches for anomaly detection. Wang and Ong (2010), for instance, combined one of the control chart techniques (MEWMA control chart) with PCA to detect structural damage using vibrational response. Loh et al. (2011) have used Singular Spectrum Analysis (SSA), a technique with a similar mathematical basis as PCA, with an autoregressive model to extract the response feature from continuous monitoring of the static deformation of a dam. More recently, Laory et al. (2013) applied Moving PCA (MPCA) in combination with four regression analysis methods for damage detection in bridges. They compared these combined methods with stand-alone methods to see which ones provided highest levels of damage detectability as well as earlier detections.

In this paper, we present the application of MPCA, which has been tested by Laory et al. (2013), to see if this method can be also useful when analyzing piezometer data from embankment dams. The main motivation is to improve piezometer data analysis by implementing statistical anomaly detection, thus reducing the subjective quality of the analyses that are commonly carried out by engineers today and the errors that come from this practice. In the next section, we present a brief theoretical explanation of the proposed detection method, followed by a description of the case study dam as well as the application and results.

MOVING PRINCIPAL COMPONENT ANALYSIS (MPCA) AND ROBUST REGRESSION ANALYSIS (RRA)

By decomposing a data matrix into a number of independent components, PCA can identify periodic variations that are dominant in the data. While PCA is often applied to the whole dataset, it can also be applied to a subset, or a window, of

the dataset. MPCA performs PCA by sliding this window, so that any change in the first several principal components over time can be detected.

More formally, consider a data matrix, $T \in \mathbb{R}^{N \times M}$, whose M columns are individual time-series of length N (e.g., measurements from individual piezometers) that have been normalized with respect to each column. Each entry of this matrix can be denoted by $V_i(t)$, where $i = 1, \dots, M$ and $t = 1, \dots, N$, as shown in the equation below. $V_i(t)$ would indicate the measurement of piezometer i at time t.

$$T = \begin{bmatrix} V_1(1) & V_2(1) & \cdots & V_M(1) \\ V_1(2) & V_2(2) & \cdots & V_M(2) \\ \vdots & \vdots & \cdots & \vdots \\ V_1(N) & V_2(N) & \cdots & V_M(N) \end{bmatrix}$$

To apply MPCA on T, first a sliding window of size L is applied to the matrix, to extract a sub-matrix, called $R_i(k)$ at each time value k, where $k = 1, \dots, (N - L)$. Then, a singular value decomposition (SVD) is performed on each one of the covariance matrices, $C = \frac{1}{N} R_i(k)^T \times R_i(k)$. During SVD, the matrix, T gets decomposed into matrices $U, S,$ and V, where $C = U * S * V^T$. The columns of U are the left singular vectors while those of V are the right singular vectors. S is a diagonal matrix with singular values along the diagonal. Since C is symmetric, the right singular vectors correspond to the eigenvectors, E_l and the diagonal elements of S corresponds to the eigenvalues, e_l, of the matrix (where $l = 1, \dots, L$).

The eigenvectors of the covariance matrix represent the directions of maximum variance, or the variance of each independent component, and the corresponding eigenvalue indicates a degree of each component's proportional variance. Thus, the most dominant patterns can be captured by the first few sets of the eigenvectors after ordering the corresponding eigenvalues in a descending order.

Now that the direction of most variability is known for each time step, the next phase is to determine whether this eigenvector changes over time, which would signal the presence of an anomaly. Robust Regression Analysis is known as a good regression technique to use in the presence of outliers. Among many types of robust regression models, we employed the method that uses iteratively reweighted least squares with a bisquare weighting function. So at the end of the proposed anomaly detection, RRA is performed to observe if any changes in the first few relevant eigenvectors from $R_i(k)$ have occurred over time. The number of eigenvectors to be monitored can be determined based on how sensitive the anomaly detection needs to be. The regression model is formed based on the normal data, and the threshold level is determined by computing the standard deviation of absolute values of the regression residuals (a difference between actual and predicted values) in the normal data. Any regression residuals that exceed this threshold are marked as anomalies.

CASE STUDY

To test the efficacy of this anomaly detection algorithm, we chose a case study dam located in the eastern part of the U.S., for which we had access to

instrumentation measurements. The dam is an earth and rock fill structure and is composed of a central core of impervious rolled fill with the upstream side slope protected by rock fill on gravel bedding. On the downstream of the core, there is a large zone of rock fill. There are a total of 26 piezometers installed and, currently, measurements are collected automatically every 4 hours.

Data validation and preparation. According to currently practiced validation criteria, any sensor measurements that fall out of the range between the top and tip elevations of the piezometers get eliminated. In addition, if there are missing data due to satellite transmission problems, these voids are filled by interpolation between the previous and proceeding readings. In addition, whenever there is no change in values among three consecutive readings, or within 12 hour-period, those readings get validated by engineers. In the application, we used the validated data that have passed the aforementioned criteria. Before September 2009, the piezometer data have been collected every 12 hours, instead of every 4 hours. To unify these two different frequencies, the time-series sampled at 4-hour intervals was down-sampled to match the 12-hour sampling period of the other time-series by selecting every fourth measurement from the 4-hour data.

APPLICATIONS AND RESULTS

The piezometer is the most common instrument used to measure water level on dams (Crum 2011). Besides the reservoir pool, piezometers are influenced by many factors such as precipitation, tail water, pressure, temperature, etc. However, since the main influencing variable to the collected piezometer data is the pool level, we applied MPCA to this pair of highly correlated variables (i.e., pool and one of the piezometers installed in the case study dam). The Pearson's correlation coefficient between the two variables is 0.983.

Other than minor seepage, which is common in embankment dams, there have been neither major structural problems nor serious turbid water effects in the dam. Thus, to test the ability of our approach to detect anomalies as well as to reduce any bias of where the anomalies are introduced, we decided to simulate anomalies to a number of time periods with a constant interval. By setting the first two years of the data as a normal condition of the dam, 4 months of anomalous data were introduced for each subsequent time period in each experiment. Given the length of the piezometer data, we could experiment with 6 unique abnormal time periods without overlaps. To simulate the anomalies, the original piezometer data during the chosen "abnormal" periods were randomly reordered in time. We chose this approach, instead of artificially generating new data, so that the resulting data were kept within the piezometer ranges of the original periods and the primary effect would be a de-correlation between the pool level and this particular piezometer during that period. Even though we have not characterized real anomalies, this simulated anomaly would represent a problem when the piezometers are not responding to the pool levels, which may occur due to serious seepage problems and/or when the piezometers themselves are malfunctioning. To make sure the randomly reordered datasets are de-correlated compared to the original dataset, we only tested on the reordered datasets that have lower correlations than the original one.

The time series data shown in Figure 2 corresponds to one of the 6 experiments we have tested on. The piezometric elevation corresponds to the readings from one of the piezometers installed. A dotted rectangle indicates the period where the anomalies were added to the piezometer data. While the normal piezometer data remain highly correlated with the pool data, the abnormal piezometer data no longer respond well to the pool levels. Both readings shown in Figure 2 indicate non-normalized data.
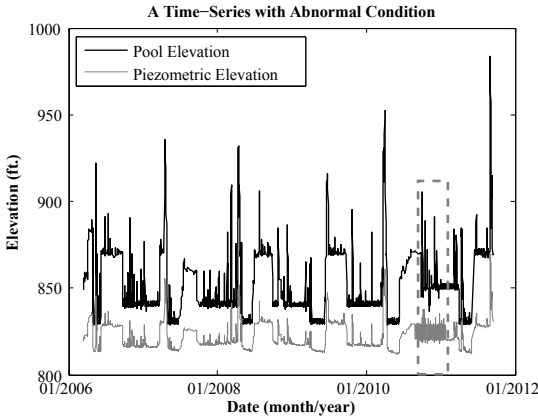
Figure 2. Time-series of the pool and the piezometric elevations with anomalies (dotted rectangle)

When applying MPCA, a window size of 730, which corresponds to a year, was used to capture a periodic behavior of the dam. In this experiment, we observed the changes in the first eigenvectors only. After computing eigenvectors through MPCA, Robust Regression Analysis was performed. The regression residuals were calculated based on the normal data, and 13 different standard deviations (from 3 to 15) of the regression residuals were cross-validated over the 6 experiments to see which standard deviation can detect the anomalies most accurately. The Receiver Operating Characteristic (ROC) curve of the 13 standard deviations is shown in Figure 3. When the distance between the best possible detection point at (0,1) and each data point in Figure 3 was computed, the standard deviation of 6 had the shortest distance. The true positive rate (TPR) over the six experiments when ± 6 standard deviations were used was 0.82 while the false positive rate (FPR) was 0.12. This high TPR and the low FPR validate that the tested method can detect anomalies in the piezometer data successfully with an average accuracy of 0.86, which was obtained by taking the ratio of sum of true positives and negatives to the total. The contingency tables for all of the 6 experiments using the ± 6 standard deviations are also included in Table 1. As a background check, a dataset of no anomaly was tested also using ± 6 standard deviations, and 1238 false positives and 1957 true negatives were found.
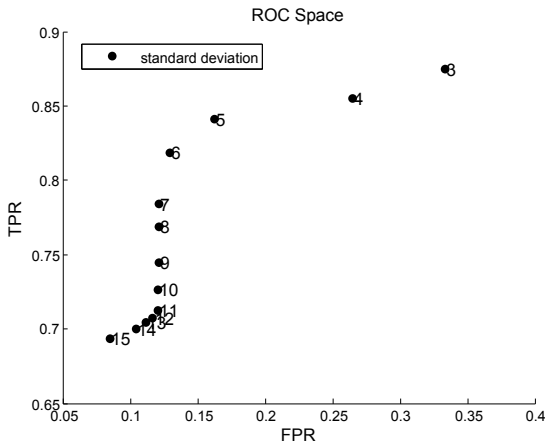
Figure 3. The ROC space from the 13 standard deviations

Table 1. Contingency Tables for the 6 Experiments

Exp.1		Exp. 2		Exp. 3		Exp. 4		Exp. 5		Exp. 6	
TP= 939	FP= 304	860	304	338	304	901	304	934	304	839	195
FN= 41	TN= 1911	120	1911	642	1911	79	1863	46	1911	141	2020

CONCLUSION

In an effort to implement a quantitative and robust approach to monitor the performance of embankment dams based on the piezometer data, we have applied MPCA and RRA as anomaly detection. To test anomaly detectability, 6 different anomalous datasets were introduced subsequently after the period of the normal data. Then, several standard deviations of the regression residuals were cross-validated to find a proper threshold level. The results of the 6 experiments presented a high true positive rate. Thus, the proposed anomaly detection method has the potential as a promising data-driven method to analyze piezometer measurements.

When a contingency table for each experiment was generated, variations among the numbers of true positives, true negatives, false positives and false negatives could be observed. This result is due to randomly generated anomalies that were introduced in each experiment. Thus, instead of shuffling the original data, it would be also interesting to simulate anomalies that are caused by other scenarios, such as change in trends, movements in materials due to hydraulic loadings, high pool events, etc.   In addition, other than observing changes in the relationship between pool and piezometer data, applying MPCA to a group of piezometers that

have similar characteristics (e.g., response times, soil layers, etc.), would be useful. If multiple piezometers can be analyzed at the same time, engineers would not need to evaluate every single piezometer with pool data.

REFERENCES

ASCE. (2000). *Guidelines for Instrumentation and Measurements for Monitoring Dam Performance*. ASCE Publications.

ASCE. (2009). *2009 Infrastructure Fact Sheet: Facts about DAMS*. 2009 Report Card for America's Infrastructure, 15–23.

ASDSO, and FEMA. (2012). "Living With Dams: Know Your Risks." NDSP-National Dam Safety Program.

Betti, R., Lus, H., Franco, G., Imbimbo, M., and Yu, J. (2006). "Identification of structural damage using dynamic input-output measurements." *Proc. SPIE 6174, Smart Structures and Materials 2006: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, 61741K*, 61741K–61741K.

Crum, D. (2011). "Rogue Piezometers." *31st Annual USSD Conference*, 21st Century Dam Design - Advances and Adaptations, 1591–1602.

Gall, E. (2007). "Correlation Plot-A Modern Tool in Dam Safety Monitoring." *27th Annual USSD Conference*, Modernization and Optimization of Existing Dams and Reservoirs, 537–551.

Laory, I., Trinh, T., Posenato, D., and Smith, I. (2013). "Combined Model-Free Data-Interpretation Methodologies for Damage Detection During Continuous Monitoring of Structures." *Journal of Computing in Civil Engineering*, (ja).

Loh, C.-H., Chen, C.-H., and Hsu, T.-Y. (2011). "Application of Advanced Statistical Methods for Extracting Long-Term Trends in Static Monitoring Data from an Arch Dam." *Structural Health Monitoring*.

Tibaduiza, D. A., Mujica, L. E., and Rodellar, J. (2012). "Damage classification in structural health monitoring using principal component analysis and self-organizing maps." *Structural Control and Health Monitoring*, n/a–n/a.

USSD. (2010). "The Aging of Embankment Dams." (The Aging of Embankment Dams), 1–11.

Wang, Z., and Ong, K. C. G. (2010). "Multivariate Statistical Approach to Structural Damage Detection." *Journal of Engineering Mechanics*, 136(1), 12–22.

Worden, K., and Manson, G. (2007). "The application of machine learning to structural health monitoring." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 515–537.

YAN, A.-M., and Golinval, J.-C. (2006). "An Enhanced Principal Component Analysis for Structural Health Monitoring." *Structural Health Monitoring 2006: Proceedings of the Third European Workshop*, 555–562.

Ying, Y. (2012). "A Data-Driven Framework for Ultrasonic Structural Health Monitoring of Pipes." *Dissertations*.

Yu, L., Zhu, J., and Chen, L. (2010). "Parametric study on PCA-based algorithm for structural health monitoring." *Prognostics and Health Management Conference, 2010. PHM '10.*, 1 –6.

A novel data utilization and control strategy for wireless structural control systems with TDMA network

Z. Sun[1], B. Li[2], S. J. Dyke[3] and C. Lu[4]

[1]School of Mechanical Engineering, Purdue University, West Lafayette, IN 47906, USA; email: sun152@purdue.edu
[2]Department of Computer Science and Engineering, Washington University in St. Louis, MO 63130, USA; email: boli@seas.wustl.edu
[3]School of Mechanical Engineering, Purdue University, West Lafayette, IN 47906, USA; email: sdyke@purdue.edu
[4]Department of Computer Science and Engineering, Washington University in St. Louis, MO 63130, USA; email: lu@cse.wustl.edu

ABSTRACT

Wireless control systems have advantages of flexible installation, rapid deployment, low maintenance cost and low power consumption compared to wired control system. The attractive features of wireless system as well as the rapid development of wireless sensor hardware, software and middleware make wireless control a promising field. However, there are also issues associated with wireless control systems such as network induced delay and data loss. These effects can seriously affect the control performances and may even cause instabilities in the systems. In this paper, a novel data utilization and control strategy is proposed for wireless control system with Time Division Multiple Access (TDMA) network. The proposed strategy utilizes the maximum throughput of TDMA network and is evaluated using a wireless control and monitoring simulator (WCMS) for benchmark active mass driver (AMD) control system. The WCMS consists of realistic structural control components and network components. The structural control components are built in Simulink® (Matlab®) and network components are built in wireless network simulation tool – TOSSIM. The proposed strategy is shown to be effective in controlling the AMD system with TDMA network induced delay and data loss impact.

INTRODUCTION

In the past few decades, structural control systems have been given more attention and considerable research have been done for structural vibration control under wind loading or seismic excitation (Spencer & Nagarajaiah, 2003). In a traditional structural control system, a large amount of cables are deployed for communication among sensors, controllers and actuators. In such a system, implementation of wired sensors is quite expensive and laborious. To reduce installation time and maintenance cost, wireless structural control systems are

proposed as an alternative for real-time structural control. Wireless sensors are capable of taking measurements from structure, communication with each other and sending collected measurements back to controller for structural control purpose. The attractive features of wireless system as well as the rapid development of wireless sensor hardware, software and middleware make wireless control a promising field.

Related research in wireless structural control include the following: Researchers from Georgia Institute of Technology deployed wireless sensor networks (WSNs) on a three-story full scale lab structure. Centralized, partially decentralized and fully decentralized wireless control architectures are designed and implemented with semi-active control devices (Wang, *et al.*, 2007a). Researchers from University of Illinois developed TinyOS-based real-time data acquisition framework on the imote2 platform for wireless structural control. The communication and processing protocol allows for near-real-time sensing of 108 channels across 27 nodes with minimal data loss. 3-sample buffered approach and 9 sample buffered approach are proposed for large network size (Linderman, *et al.*, 2012). Researchers from University of Michigan implemented partially decentralized control architecture on a six-story lab structure. Onboard computation abilities of the smart wireless sensors are utilized for state estimation and desired control force calculation in the experiment. The wireless control system has been shown to be effective in controlling the multi-story structure (Swartz & Lynch, 2009).

In this paper, centralized wireless structural control architecture with TDMA network is investigated. A data utilization strategy is proposed for centralized architecture to account TDMA delay and data loss effect to improve the structural control performance. The data utilization strategy is incorporated with optimal time-delay control algorithm for constant time delay system. The proposed strategy is evaluated using WCMS developed with benchmark active mass driver (AMD) control system.

WIRELESS DATA UTALIZATION STRATEGY FOR TDMA NETWORK

TDMA wireless communication protocol allows multiple sensors to share the same radio frequency channel by assigning a different time slot to each sensor. Each sensor (leaf node) uses its dedicated time slot to transmit data to base station (gateway node). Existing centralized wireless control approaches wait for all sensor data to arrive at the controller to calculate control force and the time interval for updating control force is restricted by the network size and the time slot size for each sensor. For ideal TDMA network without data loss, the time interval Δt for updating control force is $\sum_{i=1}^{n} h_i * \Delta$ (*n*: total sensor number, h_i: number of hops for *i*th sensor, Δ: TDMA slot). For multi-threaded WSNs, the sampling rate and transmission rate in the network are independent from each other and sampling rate is generally faster than network transmission rate.

For our proposed data utilization strategy 1 (P1), data aggregation is used to combine Δt (one round of TDMA transmission) /*dt* (sampling interval) samples into one package (Let $\Delta t/dt = k$ samples). Through this approach, the data used for control has a constant delay of $2\Delta t$ as summarized in P1. However, the time interval for

control is the same as the sampling interval. Control algorithms for constant delayed system can be used for this case. For the proposed data utilization strategy 2 (P2), data aggregation is used to combine $2k$ samples into one package. Packages between every P_{i-1} and P_i have k overlapping samples, so each sample has twice as many chances to get to the controller. Data loss rate is reduced in this design. In this approach, the time interval for control is the same as the sampling interval, and the data for control has a constant delay of $3\Delta t$. Following the pattern of P2, P3 can be developed to further reduce the data loss rate. The aggregated packet size has a limit of 117 bytes for IEEE 802.15.4 protocol. Assuming 16-bit sensor data with associated 16-bit time stamp and an 8-bit sensor ID, one package can maximum include 28 samples of data.

Proposed Data Utilization Strategy 1 (P1)

1: Initialize a WSN with n leaf node sensors;
2: Start sensing at a constant sampling interval dt;
3: **if** k samples are collected, **then**
    Aggregate k samples into one package, such that samples from step
    $(k*i+1)$ to $k*(i+1)$ aggregated at $(i+1)$ th package;
4:    Send $(i+1)$ th package from sensor j (j= 1 to n) to base station through
    TDMA network;
5:    **if** $(i+1)$ th package obtained from all sensors after one round of TDMA
      transmission, **then**
6:      Control at time step $k*(i+2)+1$ to $k*(i+3)$ with data from step $(k*i+1)$ to
      $k*(i+1)$, respectively. Optimal time-delay control algorithm for constant
      $2k$ steps delayed system is applied;
7:    **else if** package loss happens at $\forall$ sensor j
      Control at step $k*(i+2)+1$ to $k*(i+3)$ with data used at step $k*(i+2)$;
8:    **end if**
9:    **if t ≤ Final time**
10:     Update $i = i+1$; Go to step 3;
11: **end if**
12: **end if**

Proposed Data Utilization Strategy 2 (P2)

1: Initialize a WSN with n leaf node sensors;
2: Start sensing at a constant sampling interval dt;
3: **if** $2k$ samples are collected, **then**
    Aggregate $2k$ samples into one package, such that samples from step
    $(k*i+1)$ to $k*(i+2)$ aggregated at $(i+1)$ th package;
4:    Send $(i+1)$ th package from sensor j (j= 1 to n) to base station through
    TDMA network;
5:    **if** $(i+1)$ th package obtained from all sensors after one round of TDMA
      transmission, **then**
6:      Control at time step $k*(i+3)+1$ to $k*(i+4)$ with data from step $(k*i+1)$ to
      $k*(i+1)$, respectively. Optimal time-delay control algorithm for constant
      $3k$ steps delayed system is applied;

7: **else if** package loss happens at ∀ sensor *j*
  Control at step *k*(i+3)*+1 to *k*(i+4)* with data used at step *k*(i+3)*;
8: **end if**
9: **if t ≤ Final time**
10: Update *i = i*+1; Go to step 3;
11: **end if**
12: **end if**

WIRELESS STRUCTURAL CONTROL SIMULATOR

  To realistically simulate wireless control experiments, a WCMS has been proposed and established. There are three main components for the simulator which are structural control component built in Simulink®, intermediate layer built in Python and wireless network component built in TOSSIM as shown in Figure 1. Users can define the seismic input to the structure, routing algorithms, scheduling algorithms, wireless noise traces and wireless link quality values for simulating wireless network. More detailed descriptions about the simulator are provided in Sun, *et al.* (2012). For this case study, the simulator developed for benchmark AMD model is used (Figure. 2) to compare different data utilization pattern working with optimal time-delay controller.
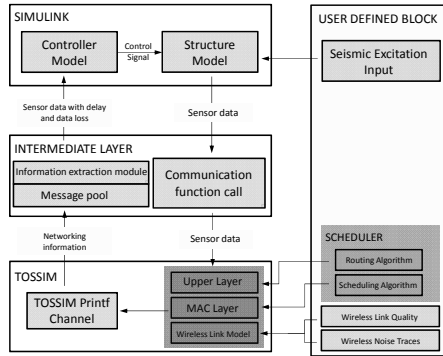
Figure 1. Components of wireless control simulator
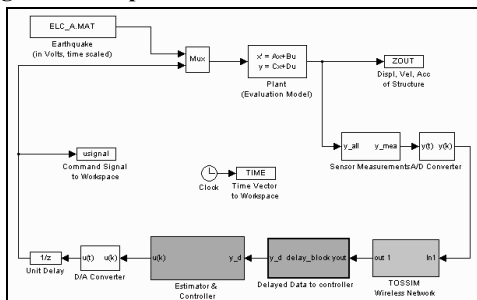
Figure 2. Simulink model for the AMD wireless control system

CONTROLLER DESIGN

For a constant time delayed system with delayed control force, the discrete system can be represented as following:

$$z[k+1] = Az[k] + Bp_d[k-l] + Ew[k] \tag{1}$$

$$y[k] = Cz[k] + Dp_d[k-l] + Fw[k] \tag{2}$$

where $z[k]$ is the discrete-time system states at k^{th} step, $p_d[k-l]$ is the $k-l$ step delayed control force, $w[k]$ is the external earthquake loading, $y[k]$ is the measured outputs at k^{th} step, A, B, C and D are system matrices, E, F are the location matrices of external loading.

For optimal time-delay control design, Eqn. 3 is used because the external earthquake loading $w[k]$ is not involved in the control algorithm derivation. In Eqn. 4 the states $z[k-l]$ are estimated from delayed measurements $y[k-l]$ due to wireless network. Kalman filter is used for the estimation.

$$z[k+1] = Az[k] + Bp_d[k-l] \tag{3}$$

$$p_d[k-l] = Gz[k-l] \tag{4}$$

$$J\mid_{p_d} = \sum_{k=l}^{\infty} \left(z_d^T[k]Qz_d[k] + p_d^T[k-l]Rp_d[k-l] \right) \tag{5}$$

where G is the optimal time-delay control gain obtained using heuristic algorithm solving coupled nonlinear matrix equations (Chung, *et al.*, 1995; Wang, *et al.*, 2007b). J is the cost function to be minimized by selecting the optimal trajectory of control force p_d. Here Q is C^TC, R is D^TD.

CASE STUDY: CONTROL OF BENCHMARK AMD BUILDING MODEL

The benchmark AMD model used in this case study is based on an actively controlled, three-story building considered in Dyke, *et al.* (1996). The test structure is designed to be a scale model of the prototype building and is subject to one-dimensional ground motion. There is a time scale factor of 0.2, making the natural frequencies of the model approximately five times those of the prototype. We also adopt this time scale for TDMA time slot in our simulation. One TDMA slot equals two sample steps in the simulation. The first three modes of the model structural system are at 5.81 Hz, 17.68 Hz and 28.53 Hz (Spencer, *et al.*, 1998). Four acceleration measurements from 1-3 floors and the active mass driver, respectively, are used for feedback control.

For the WCMS developed with this model, we adopt the El Centro Earthquake for the structure; fixed multi-hop routes obtained using shortest path routing algorithm. Wireless noise traces and link quality values are measured from a 3 story building in Washington University with TelosB sensors. Each TDMA time slot is assumed to be 10 msec (Han, *et al.*, 2011) (remote processing on leaf node, sending data and local processing at base station are all included in this 10 msec). Sensor

measurements from 1st floor need two hops to transmit to the base station and the other 3 sensors use one hop. The design for P1, P2 without data loss are shown in Table 3 and Table 4. A comparison of proposed strategies P1 and P2 with the general approach of centralized TDMA control (G1) is provided in Table 5.

Table 3. Proposed Data Utilization Pattern 1 (P1)

Sample Step	TDMA Transmission	Data available at controller	Data used for control
i+1	$P_{i,1}$	$P_{i-1,1234}$	$S_{i-19,1234}$
i+2	$P_{i,1}$	$P_{i-1,1234}$	$S_{i-18,1234}$
i+3	$P_{i,2}$	$P_{i-1,1234}, P_{i,1}$	$S_{i-17,1234}$
i+4	$P_{i,2}$	$P_{i-1,1234}, P_{i,1}$	$S_{i-16,1234}$
i+5	$P_{i,3}$	$P_{i-1,1234}, P_{i,12}$	$S_{i-15,1234}$
i+6	$P_{i,3}$	$P_{i-1,1234}, P_{i,12}$	$S_{i-14,1234}$
i+7	$P_{i,4}$	$P_{i-1,1234}, P_{i,123}$	$S_{i-13,1234}$
i+8	$P_{i,4}$	$P_{i-1,1234}, P_{i,123}$	$S_{i-12,1234}$
i+9	$P_{i,4}$	$P_{i-1,1234}, P_{i,123}$	$S_{i-11,1234}$
i+10	$P_{i,4}$	$P_{i-1,1234}, P_{i,123}$	$S_{i-10,1234}$
i+11	$P_{i+1,1}$	$P_{i-1,1234}, P_{i,1234}$	$S_{i-9,1234}$

$S_{i-19,1234}$: sensor 1, 2, 3, 4 data sampled at step i-19.
$P_{i-1,4}$: sensor 4 data package sampled from step i-19 to step i-10.
$P_{i,1}$: sensor 1 data package sampled from step i-9 to step i.

Table 4. Proposed Data Utilization Pattern 2 (P2)

Sample Step	TDMA Transmission	Data available at controller	Data used for control
i+1	$B_{i,1}$	$B_{i-1,1234}$	$S_{i-29,1234}$
i+2	$B_{i,1}$	$B_{i-1,1234}$	$S_{i-28,1234}$
i+3	$B_{i,2}$	$B_{i-1,1234}, B_{i,1}$	$S_{i-27,1234}$
i+4	$B_{i,2}$	$B_{i-1,1234}, B_{i,1}$	$S_{i-26,1234}$
i+5	$B_{i,3}$	$B_{i-1,1234}, B_{i,12}$	$S_{i-25,1234}$
i+6	$B_{i,3}$	$B_{i-1,1234}, B_{i,12}$	$S_{i-24,1234}$
i+7	$B_{i,4}$	$B_{i-1,1234}, B_{i,123}$	$S_{i-23,1234}$
i+8	$B_{i,4}$	$B_{i-1,1234}, B_{i,123}$	$S_{i-22,1234}$
i+9	$B_{i,4}$	$B_{i-1,1234}, B_{i,123}$	$S_{i-21,1234}$
i+10	$B_{i,4}$	$B_{i-1,1234}, B_{i,123}$	$S_{i-20,1234}$
i+11	$B_{i+1,1}$	$B_{i-1,1234}, B_{i,1234}$	$S_{i-19,1234}$

$B_{i-1,1}$: sensor 1 data package sampled from step i-29 to step i-10.
$B_{i,1}$: sensor 1 data package sampled from step i-19 to step i.

Table 5. Comparison of Different Patterns for Case Study

Data pattern	Data delay (sample steps)	Control interval (sample steps)	Data loss reduction
G1	10	10	No
P1	20	1	No
P2	30	1	Yes

Delay traces generated in TOSSIM wireless network simulator follow probabilistic distributions based on wireless noise traces and wireless link quality values. So each run of wireless control simulation has different results. One realization of different strategies is shown in Figure. 3. One hundred realizations of simulations have been conducted to compare three data patterns with optimal time-delay control as in Table 6. The G1 case is unstable due to large time interval for performing control and the fast dynamics of the system. Proposed data utilization P1 and P2 have achieved good control performance compared to uncontrolled case. Even though with more constant delay, performances of P2 are better than P1 as data loss in wireless transmission is reduced in the design. Wired control strategy designed with the same weighting matrices Q, R as in P1 and P2, has reduced peak and RMS accelerations responses the most.
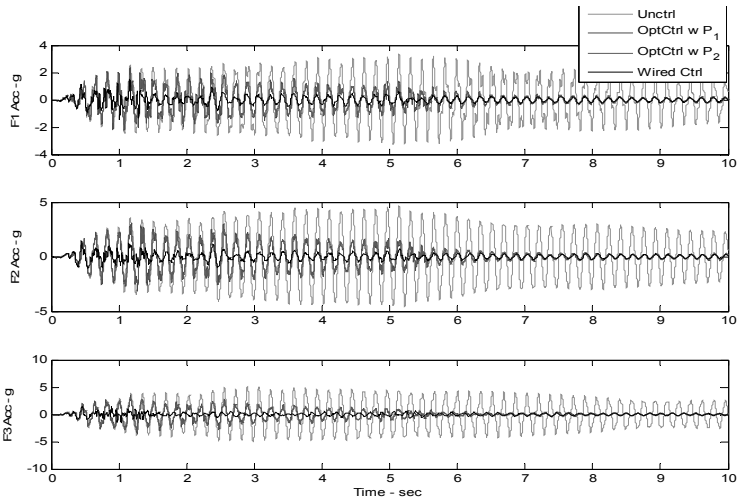
Figure 3. One realization of simulation results with El Centro Earthquake

Table 6. Peak and RMS Acc of 3 floors under El Centro Earthquake

	Unctrl (g)	G1* (g)	P1* (g)	P2* (g)	Wired Ctrl (g)
F1 Peak Acc	3.3478	-	2.3471	1.9056	1.5354
F2 Peak Acc	4.6301	-	3.0077	2.7432	1.4480
F3 Peak Acc	5.0511	-	3.4759	3.1842	1.4615
F1 RMS Acc	1.5396	-	0.6542	0.6249	0.2677
F2 RMS Acc	2.3551	-	0.9124	0.9186	0.2971
F3 RMS Acc	2.5254	-	0.8327	0.8220	0.2744

G1*: system unstable
P1*: mean value of 100 realizations of P1
P2*: mean value of 100 realizations of P2

CONCLUSIONS

Novel data utilization strategies for centralized wireless control systems with TDMA network are proposed and studied. The performance of these strategies is evaluated using the realistic wireless control simulator developed with benchmark AMD model. These strategies have been shown to be effective in vibration control for this 3-story benchmark model. Further investigations are underway to experimentally verify the control performances of proposed approaches.

ACKNOWLEDGEMENTS

REFERENCES

Chung, L. L., Lin, C. C., & Lu, K. H. (1995). Time-delay control of structures. *Earthquake Engineering & Structural Dynamics , 24* (5), 687-701.

Dyke, S. J., Spencer, B. F., Quast, P., Kaspari, D. C., & Sain, M. K. (1996). Implementation of an Active Mass Driver Using Acceleration Feedback Control. *Microcomputers in Civil Engineering* , 305-323.

Han, S., Zhu, X., Mok, A., Chen, D., & Nixon, M. (2011). Reliable and Real-time Communication in Industrial Wireless Mesh Networks. *17th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)* .

Linderman, L. E., Mechitov, K. A., & Spencer, B. F. (2012). TinyOS-based real-time wireless data acquisition framework for structural health monitoring and control. *Structural Control and Health Monitoring* .

Spencer, B. F., & Nagarajaiah, S. (2003, July). State of the Art of Structural Control. *Journal of Structural Engineering* , 845-856.

Spencer, B. F., Dyke, S., & Deoskar, H. (1998). Benchmark problems in structural control: part I-Active Mass Driver system. *Earthquake Engineering & Structural Dynamics , 27* (11), 1127-1139.

Sun, Z., Bo, L., Dyke, S., & Lu, C. (2012). Evaluation of Performances of Structural Control Benchmark Problem with Time Delays from Wireless Sensor Network. *Joint Conference of the Engineering Mechanics Institute and ASCE Joint Specialty Conference on Probabilistic Mechanics and Structural Reliability (EMI/PMC'12).*

Swartz, R., & Lynch, J. (2009). Strategic Network Utilization in a Wireless Structural Control System for Seismically Excited Structures. *Journal of Structural Engineering* , pp. 597-608.

Wang, Y., Swartz, R. A., Lynch, J. P., Law, K. H., & Loh, C.-H. (2007). Performance evaluation of decentralized wireless sensing and control in civil structures. *Proc. SPIE 6531, Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security.*

Wang, Y., Swartz, R. A., Lynch, J. P., Law, K. H., Lu, K.-C., & Loh, C.-H. (2007). Decentralized civil structural control using real-time wireless sensing and embedded computing. *Smart Structures and Systems , 3* (3), 321-340.

An Iterative Convex Optimization Procedure for Structural System Identification

Dapeng Zhu[1], Xinjun Dong[2], Yang Wang[3]

School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr NW, Atlanta, GA 30332, USA
PH (404) 894-1851; FAX (404) 894-2278; [1]email: zhudp@gatech.edu; [2]email: xdong35@gatech.edu; [3]email: yang.wang@ce.gatech.edu;

ABSTRACT

Owing to the complexity of civil structures, structural behavior predicted by finite element models (built according to design drawings) is usually different from behavior of an actual structure in the field. To improve the prediction accuracy, finite element (FE) model updating can be performed based on sensor measurement from the actual structure. Numerous FE model updating algorithms have been developed in the past few decades. However, most existing algorithms suffer computational challenges when applied to large structures. The challenges are usually because that most existing algorithms attempt to solve non-convex optimization problems. As a result, the optimization process encounters convergence difficulty and cannot guarantee global optimum. To address the issue, this paper proposes an iterative convex optimization algorithm for FE model updating. The convex attribute of the optimization problem makes the solution process tractable and efficient. To validate the proposed algorithm, numerical simulation is conducted with a plane truss structure. The proposed approach successfully updates the simulation model and shows certain advantage over a conventional FE model updating algorithm.

INTRODUCTION

In modern structural analysis, extensive efforts have been devoted to developing accurate finite element (FE) models. However, simulation results from an FE model often differ from field testing results, which may be caused by various inaccuracies in the model. For example, in actual structures, support conditions are far more complicated than ideal hinges, fixed ends, or rollers commonly used in modeling. Besides, most structural components may deteriorate over time. An FE model based on original structural properties does not accurately reflect the deteriorated structure. To obtain a more reliable model, experimental data collected from the actual structure in the field can be used to update the model, which is known as FE model updating. The updated model is expected to predict structural response

with higher fidelity. Furthermore, by tracking major property changes at individual structural components, model updating can assist in diagnosing structural damage.

During past few decades, various FE model updating methods have been developed and practically applied (Friswell and Mottershead 1995). A large amount of these methods utilize modal analysis results from field testing. Selected structural parameters are updated by forming an optimization objective function that minimizes the difference between experimental and simulated modal properties. In particular, early researchers started by minimizing difference between measured and simulated natural frequencies. For example, Zhang *et al.* (2000) proposed an eigenvalue sensitivity-based model updating approach that was applied on a scaled suspension bridge model. Salawu (1997) reviewed various model updating algorithms using natural frequencies, and concluded that changes in frequencies may not be sufficient enough for identifying system parameters. Therefore, other modal characteristics, e.g. mode shapes or modal flexibility, were investigated for model updating. For example, Moller and Friberg (1998) adopted the modal assurance criterion (MAC)-related function for updating the model of an industrial structure. FE model updating using changes in mode shapes and frequencies was applied for damage assessment of a reinforced concrete beam (Teughels *et al.* 2002). More recently, Yuen (2012) developed an efficient model updating algorithm using frequencies and mode shapes at only some selected degrees-of-freedom (DOFs) for a few modes. Jaishi and Ren (2006) proposed an objective function consisting of changes in frequencies, MAC related functions and modal flexibility for updating the model of a beam structure. Overall, most previous optimization objective functions that describe the difference in experimental and simulated modal properties are highly nonlinear and non-convex. As a result, conventional modal-based approaches suffer convergence issues, and may not guarantee global optimum.

To overcome the difficulty, this research investigates efficient model updating by formulating a convex optimization problem. The convex attribute guarantees global optimality of a solution, and makes the optimization process tractable and highly efficient (Boyd and Vandenberghe 2004; Grant and Boyd 2013; Lin *et al.* 2010). In addition, some DOFs are difficult to measure during field experiments. A modal expansion process using measurement at a limited number of DOFs is adopted in order to obtain experimental mode shapes for all DOFs. Due to modal expansion, formulation of the convex optimization problem is based upon an initial FE model. Therefore, an iterative convex optimization procedure is proposed for higher updating accuracy. After an updated model is obtained as the solution of convex optimization, the updated model is used again as an initial model to repeat the updating process till convergence. The rest of the paper is organized as follows. The formulation of iterative convex optimization procedure is presented first. Numerical validation on a plane truss model is then described. Performance of the proposed approach is compared with a conventional model updating approach. Finally, a summary and discussion are provided.

ITERATIVE CONVEX OPTIMIZATION FOR MODEL UPDATING

For a linear structural system, the system stiffness and mass matrices can be updated as:

$$\mathbf{K} = \mathbf{K}_0 + \sum_{i=1}^{n_\alpha} \alpha_i \mathbf{K}_{0,i} \qquad \mathbf{M} = \mathbf{M}_0 + \sum_{j=1}^{n_\beta} \beta_j \mathbf{M}_{0,j} \qquad (1)$$

where $\mathbf{K}_0$ and $\mathbf{M}_0$ are the initial estimated stiffness and mass matrices prior to model updating; α_i and β_j are system stiffness and mass parameters to be updated (such as elastic modulus and density of each element); n_α and n_β represent the total number of updating parameters; $\mathbf{K}_{0,i}$ and $\mathbf{M}_{0,j}$ are constant matrices that represent contributions corresponding to the updating parameters. From Eq. (1), the system matrices are affine functions of updating parameters $\boldsymbol{\alpha} = \left[\alpha_1, \alpha_2, \cdots, \alpha_{n_\alpha} \right]^{\mathrm{T}}$, and $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \cdots, \beta_{n_\beta} \right]^{\mathrm{T}}$.

In this study, a convex optimization formulation is adopted for model updating with optimization variables $\mathbf{x} = \left[\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{s} \right]^{\mathrm{T}}$:

$$\text{Minimize} \quad \max \left(s_1, s_2, \cdots, s_m \right) \qquad (2a)$$

$$\text{Subject to} \quad \left\| \left(-\omega_j^2 \mathbf{M} + \mathbf{K} \right) \boldsymbol{\psi}_j \right\| - s_j \le 0; \quad j = 1...m \qquad (2b)$$

$$\boldsymbol{\alpha}_L \le \boldsymbol{\alpha} \le \boldsymbol{\alpha}_U \qquad \boldsymbol{\beta}_L \le \boldsymbol{\beta} \le \boldsymbol{\beta}_U \qquad (2c)$$

where m denotes the number of measured modes obtained from experimental data; $\mathbf{s} = \left[s_1, s_2, \cdots, s_m \right]^{\mathrm{T}}$ includes slack variables limiting the 2-norm of eigenvalue expressions in Eq. (2b); $\boldsymbol{\alpha}_L$ and $\boldsymbol{\beta}_L$ denote the element-wise lower bounds for vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively; $\boldsymbol{\alpha}_U$ and $\boldsymbol{\beta}_U$ denote the element-wise upper bounds for vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, respectively; ω_j and $\boldsymbol{\psi}_j$ represent the j-th frequency and mode shape, obtained from experimental data. Note that the sign $"\le"$ in Eq. (2c) is overloaded to represent element-wise inequality.

It can be proved that the optimization problem in Eq. (2) satisfies the definition of a convex optimization problem (Boyd and Vandenberghe 2004; Zhu and Wang 2012) by assuming that the experimental frequencies and mode shapes (ω_j and $\boldsymbol{\psi}_j$, $j = 1,2,...m$) are constant. However, in practice, usually not all DOFs can be measured by sensors. This indicates that only incomplete mode shapes can be directly obtained from experimental data. To solve this issue, modal expansion can be performed (Kidder 1973):

$$\boldsymbol{\psi}_{j,U} = (-\mathbf{D}_{UU}^{-1} \mathbf{D}_{UM}) \boldsymbol{\psi}_{j,M} \qquad (3)$$

where subscripts M and U represent the measured and unmeasured DOFs, respectively; $\boldsymbol{\psi}_{j,M}$ and $\boldsymbol{\psi}_{j,U}$ represent the measured and unmeasured entries of the j-mode shape vector. The expansion matrix $(-\mathbf{D}_{UU}^{-1} \mathbf{D}_{UM})$ comes from the eigenvalue problem of the initial structural model:

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_{MM} & \mathbf{D}_{MU} \\ \mathbf{D}_{UM} & \mathbf{D}_{UU} \end{bmatrix} = -\omega_j^2 \mathbf{M}_0 + \mathbf{K}_0 \qquad (4)$$

As described before, the model updating process is based upon an initial FE model. The updated model can be used as an initial model again to repeat the updating process for higher accuracy. The procedures of the iterative convex optimization process for model updating are summarized as follows.

(1) Formulate the initial FE model ($\mathbf{K}_0$ and $\mathbf{M}_0$), e.g. according to design drawings;

(2) Modal expansion using experimental mode shapes measured at a limited number of DOFs (Eq. (3));

(3) Solve convex optimization problem for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\mathbf{s}$ (Eq. (2));

(4) Use optimal $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ to update $\mathbf{K}_0$ and $\mathbf{M}_0$, and return to Step (2), until the updated variables converge.

NUMERICAL EXAMPLE

To validate the proposed convex optimization approach for model updating, a plane truss model is simulated (Figure 1). The truss model has 14 nodes, and each node has two translational DOFs. Table 1 summarizes the structural properties of the model. Different elastic modulus is assigned to top chords, diagonal & vertical bars, and bottom chords. Non-ideal support conditions are considered in this example. Vertical and horizontal springs are allocated at the left support to simulate a non-ideal hinge, while a vertical spring is allocated at the right support to simulate a non-ideal roller. Figure 2 shows the first six natural frequencies and mode shapes of the structure. For simplicity, these modal properties are directly used as the experimental results for model updating.
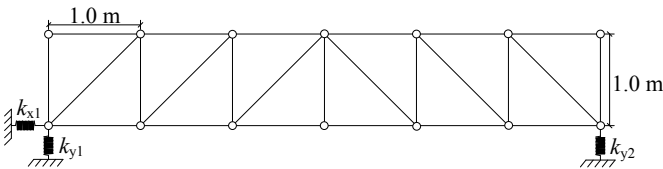
Figure 1. Plane Truss Structure

Table 1. Structural Properties

Property		Value
Elastic modulus (N/m^2)	Top chords	2.2×10^{11}
	Diagonal & vertical bars	2.1×10^{11}
	Bottom chords	1.9×10^{11}
Density (kg/m^3)		7,849
Cross-section area (m^2)		8×10^{-5}
Spring k_{y1} (N/m)		7×10^{6}
Spring k_{x1} (N/m)		2×10^{6}
Spring k_{y2} (N/m)		5×10^{6}
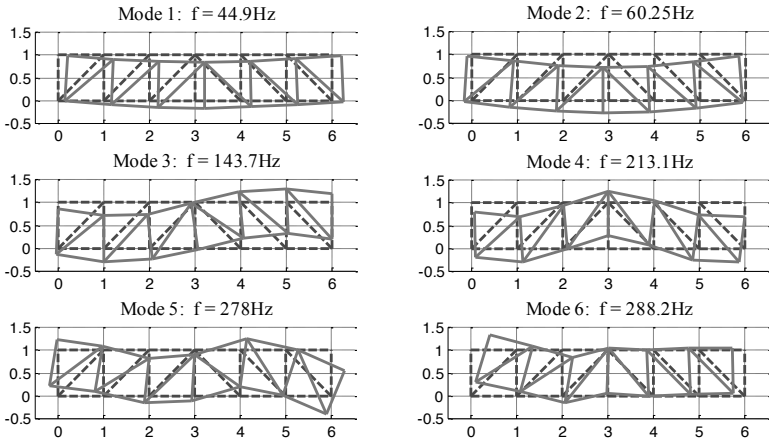
Figure 2. Modal Properties of the Plane Truss Structure

For comparison, a conventional model updating approach is also studied (Jaishi and Ren 2006). The conventional model updating formulation aims to minimize the difference between experimental and analytical natural frequencies and mode shapes. The optimization problem has optimization variables $\mathbf{x} = [\boldsymbol{\alpha}, \boldsymbol{\beta}]^{\mathrm{T}}$, which corresponds to stiffness and mass updating parameters.

$$\text{Minimize} \quad \sum_{i=1}^{m} \left(\frac{f_{\mathrm{FE},i} - f_{\exp,i}}{f_{\exp,i}} \right)^2 + \sum_{i=1}^{m} \left(\frac{1 - \sqrt{\mathrm{MAC}_i}}{\sqrt{\mathrm{MAC}_i}} \right)^2 \tag{5a}$$

$$\text{Subject to} \quad \boldsymbol{\alpha}_L \le \boldsymbol{\alpha} \le \boldsymbol{\alpha}_U \qquad \boldsymbol{\beta}_L \le \boldsymbol{\beta} \le \boldsymbol{\beta}_U \tag{5b}$$

Here m denotes the number of measured modes; $f_{\mathrm{FE},i}$ and $f_{\exp,i}$ represent the analytical (from FE model) and experimental natural frequencies, respectively; MAC_i represents the modal assurance criterion evaluating the difference between the i-th analytical and experimental mode shapes.

A nonlinear least square optimization solver, 'lsqnonlin' in MATLAB toolbox (MathWorks Inc. 2005), is adopted to numerically solve the conventional model updating problem. The optimization solver seeks a minimum of the objective function in Eq. (5a) through Levenberg-Marquardt algorithm (Moré 1978), which uses a search direction interpolated between the Gauss-Newton direction and the steepest descent direction.

Two measurement cases are studied. Case 1 assumes all DOFs are measured. Because modal expansion is not needed, model updating through the proposed approach is achieved by solving the convex optimization problem (Eq. (2)) only once, i.e. without iteration. Case 2 assumes partial DOFs are measured. Because modal expansion is needed, the iterative convex optimization procedure is performed. For comparison, model updating of both cases are also performed using conventional

formulation in Eq. (5). In this preliminary study, it is assumed to have perfect knowledge on structural mass. Therefore, only stiffness parameters $\boldsymbol{\alpha}$ are updated. After model updating, the root mean square (RMS) of the relative difference between optimal and actual parameters (Table 1) is calculated to evaluate the updating performance.

$$RMS = \sqrt{\frac{1}{n_\alpha} \sum_{i=1}^{n_\alpha} \left(\frac{p_{opt,i} - p_{act,i}}{p_{act,i}} \right)^2} \tag{6}$$

Here n_α denotes the number of stiffness parameters being updated; $p_{opt,i}$ and $p_{act,i}$ represent the optimal value after updating and the actual value, respectively.

Case 1   In this case, since all DOFs are measured, modal expansion is not necessary. Table 2 lists all the updating parameters, including elastic moduli and support spring stiffnesses. The initial values of the updating variables are randomly assigned to be different from actual values. Both the proposed approach and the conventional model updating approach are performed. For each approach, the updating is performed assuming different numbers of measured modes (i.e. modes corresponding to the 1, 2, or 6 lowest natural frequencies) are available. The initial and updated parameter values are summarized in Table 2.
        Shown by Table 2, the conventional formulation (Eq. (5)) can achieve optimal solutions when two or more modes are available for model updating. However, when only one mode is available, results of the conventional formulation have 8.44% RMS error from actual values. The proposed convex optimization algorithm can always achieve optimal values, even when only one mode is available. The corresponding

Table 2. Model Updating Results using Complete Measurements

Updating parameter		Elastic modulus (10^{11} N/m^2)			k_{y1} (10^6 N/m)	k_{y2} (10^6 N/m)	k_{x1} (10^6 N/m)	RMS error (%)
		Top chords	Diag. & Vert.	Bottom chords				
Initial value		2.000	2.000	2.000	6.000	6.000	6.000	82.4
Proposed approach	1 mode	2.200	2.100	1.900	7.000	5.000	2.000	0.00
	2 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
	6 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
Conventional approach	1 mode	2.318	2.373	1.810	6.612	5.659	2.033	8.44
	2 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
	6 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
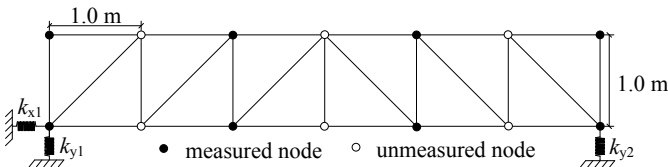
Figure 3. Measurement Configuration for Case 2

Table 3. Model Updating Results using Incomplete Measurements

Updating parameter		Elastic modulus $(10^{11} \, \text{N/m}^2)$			k_{y1} $(10^6 \, \text{N/m})$	k_{y2} $(10^6 \, \text{N/m})$	k_{x1} $(10^6 \, \text{N/m})$	RMS error (%)
		Top chords	Diag. & Vert.	Bottom chords				
Initial value		2.000	2.000	2.000	6.000	6.000	6.000	82.4
Proposed approach	1 mode	2.195	2.100	1.900	6.998	4.999	2.000	0.10
	2 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
	6 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
Conventional approach	1 mode	2.292	2.329	1.825	6.585	5.410	2.028	6.53
	2 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00
	6 modes	2.200	2.100	1.900	7.000	5.000	2.000	0.00

RMS errors are zero regardless how many modes are available.

<u>Case 2</u>　　　In this case, it is assumed only some nodes are measured by sensors. Modes shapes directly extracted from sensor data are only available at the measure DOFs. Figure 3 illustrates the measured nodes with solid dots. Both horizontal and vertical DOFs are measured at every solid dot.　Similar to Case 1, both the conventional updating approach and the iterative convex optimization approach are applied. Table 3 summarizes the updating results of both approaches using different numbers of measured modes. It should be noted that the modal expansion (Eq. (3)) in the iterative convex method gives better estimation for modes at lower frequencies, but may introduce significant errors when expanding higher modes. Therefore, when multiple modes are available, it is decided to only use lower modes at the first few iterations for preliminary updating.　Higher modes are then used to refine the updating results. The modal expansion process is not needed in conventional approach (Eq. (5)).

　　　　Table 3 demonstrates that the two updating approaches have similar performance as in previous Case 1.　The conventional updating approach gives accurate solution, when two or more modes are available, but does not provide satisfactory results when only one mode is available. The iterative convex optimization approach works well when different numbers of modes are available. This advantage of the iterative convex optimization is promising, because in field testing, measurable structural vibration energy mainly occupies low frequency range. Modal properties of lower modes are easier to obtain reliably than higher modes.

CONCLUSION

　　　　This research investigates FE model updating through the formulation of a convex optimization problem. The convex attribute makes the solution process tractable and efficient.　In order to apply the updating when only a limited number of DOFs are measured, modal expansion is performed to obtain the complete mode shapes. As required by modal expansion process, formulation of the convex optimization problem requires an initial FE model. An iterative optimization procedure is proposed to incorporate modal expansion at every iterative step, and to

achieve higher updating accuracy. Numerical simulation is conducted to validate the proposed approach, and to compare the performance with a conventional approach minimizing the difference between analytical and experimental modal properties. It is shown that the proposed approach may give better performance when very limited numbers of modes are available. Nevertheless, more extensive analytical and numerical studies are needed on the convergence, accuracy, and computational efficiency of the proposed method.

REFERENCES

Boyd, S. P., and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge University Press, Cambridge, UK ; New York.

Friswell, M. I., and Mottershead, J. E. (1995). *Finite element model updating in structural dynamics*, Kluwer Academic Publishers, Dordrecht; Boston.

Grant, M., and Boyd, S. (2013). CVX: MATLAB software for disciplined convex programming, version 2.0. http://stanford.edu/~boyd/cvx.

Jaishi, B., and Ren, W. X. (2006). "Damage detection by finite element model updating using modal flexibility residual." *Journal of Sound and Vibration*, 290(1-2), 369-387.

Kidder, R. L. (1973). "Reduction of structural frequency equations." *AIAA Journal*, 11(6), 892-892.

Lin, M. M., Dong, B., and Chu, M. T. (2010). "Semi-definite programming techniques for structured quadratic inverse eigenvalue problems." *Numerical Algorithms*, 53(4), 419-437.

MathWorks Inc. (2005). *Control System Toolbox : for Uses with MATLAB® : Getting Started*, Version 6., MathWorks Inc., Natick, MA.

Moller, P. W., and Friberg, O. (1998). "Updating large finite element models in structural dynamics." *AIAA Journal*, 36(10), 1861-1868.

Moré, J. (1978). "The Levenberg-Marquardt algorithm: Implementation and theory." Numerical Analysis, G. A. Watson, ed., Springer Berlin Heidelberg, 105-116.

Salawu, O. S. (1997). "Detection of structural damage through changes in frequency: A review." *Engineering Structures*, 19(9), 718-723.

Teughels, A., Maeck, J., and De Roeck, G. (2002). "Damage assessment by FE model updating using damage functions." *Computers & Structures*, 80(25), 1869-1879.

Yuen, K. V. (2012). "Updating large models for mechanical systems using incomplete modal measurement." *Mechanical Systems and Signal Processing*, 28, 297-308.

Zhang, Q. W., Chang, C. C., and Chang, T. Y. P. (2000). "Finite element model updating for structures with parametric constraints." *Earthquake Engineering & Structural Dynamics*, 29(7), 927-944.

Zhu, D., and Wang, Y. "Substructure model updating through iterative convex optimization." *Proceeding of the ASME 2012 Conference on Smart Materials, Adaptive Structures and Intelligent Systems (SMASIS 2012)*, Stone Mountain, GA, USA.

Algorithmic and Computing Technologies for Health Assessment of Real Structures in the Presence of Nonlinearity and Uncertainty

Ajoy Kumar Das[1], Abdullah Al-Hussein[1] and Achintya Haldar[2*]

[1]Doctoral Student, Department of Civil Engineering and Engineering Mechanics, University of Arizona, P.O. Box 210072, Tucson, AZ 85721, USA; PH +1 (520) 621-2266; FAX: (520) 621-2550; Email: akdas@email.arizona.edu; Email: abdullaa@email.arizona.edu
[2]Professor, Department of Civil Engineering and Engineering Mechanics, University of Arizona, P.O. Box 210072, Tucson, AZ 85721, USA; PH +1 (520) 621-2142; FAX: (520) 621-2550; Email: haldar@u.arizona.edu

ABSTRACT

The research team at the University of Arizona proposed several novel structural health assessment (SHA) algorithms. Structures are represented by finite elements (FE) and the health is assessed by identifying the stiffness parameters of all the elements and comparing them with expected values, or with previous values, or observing differences between similar elements. They can identify the location and severity of defect and exact location within a defective element. These algorithms use several system identification (SI) based concepts with different levels of sophistications. They do not require excitation information and can assess the health of large structural systems using only limited noise-contaminated acceleration time-histories measured at a small part of a structure. They are widely available in the literature. However, algorithmic and computational rigors of them are generally not presented in technical papers due to severe page limitation. Some of them are briefly presented in this paper without discussing the specific algorithms.

INTRODUCTION

Health assessment and monitoring (HAM) of infrastructures has become one of the important research areas. Globally an estimate of 2% of gross domestic product (GDP), about U.S. $960 billion, is spent annually on infrastructure investment and maintenance. Based on the Report Card of 2005 for Americas Infrastructure, it was estimated that over $1.6 trillion over a five-year period was necessary to bring the nation's infrastructure to a good condition.

This enormous challenge has attracted multidisciplinary research interest. One line of approach has been to extend the life of infrastructures by locating defects and evaluating their severity at the local element level and taking the appropriate remedial actions without exposing public to out-of-ordinary risk. The authors have been

working on several finite element-based system-identification (SI) algorithms for HAM using measured acceleration time-histories. The basic concept is as follows. During an inspection, several accelerometers will be placed at pre-determined strategically located FE node points in a large structure and acceleration time-histories will be measured for a very small period of time may be less than a second, to avoid contamination from other excitation sources. The task is to extract signature hidden in the response information to assess structural health at the element level.

The research team at the University of Arizona proposed several algorithms. In all these algorithms, using only measured acceleration time-histories, the stiffness parameters of all the elements in the FE representation are identified. The structural health is assessed by comparing the identified stiffness parameters with expected stiffness parameters from the design drawings or compared with previous values if inspected previously, or changes between two similar elements. This way, not only the location but also the severity of the defects can be established. Once a defective element is identified, the exact defect spot within the element can be established by refining the FE model

There are several unique algorithmic and computational features of these methods. Since excitation time-histories are rarely available during the SI process, it will be extremely desirable if a system can be identified without excitation information. For large realistic systems, it may be practically impossible to measure responses at all dynamic degrees of freedom (DDOFs); they may be measured at a small part of the system. Furthermore, all measured responses are expected to be noise-contaminated even when measured by smart sensors. The responses can be highly nonlinear depending on the severity of damages and/or excitation. The challenge is to identify a large system using only limited noisy responses measured at a small part of the system in the presence of nonlinearity and uncertainty.

It is not possible to discuss different SI-based procedures developed by the team (Wang and Haldar, 1994, 1997; Ling and Haldar, 2004; Katkhuda et al., 2005; Katkhuda and Haldar, 2008; Das and Haldar, 2012a, b). However, algorithmic and computational features used in developing these methods were not clearly stated in these publications due to lack of space. Some of these features are specifically addressed in this paper.

ADVANCED ALGORITHMIC AND COMPUTATIONAL FEATURES

System Identification with Unknown Input

A typical SI-based algorithm using dynamic response information will have three essential components: (1) excitation force, (2) the system to be identified represented by FEs, and (3) measured response information. It will be reasonably straight forward to identify the system if excitation and response information is available. Excitation information may not be available after a natural disaster or it may be so much noise-contaminated that the use of any SI-based algorithm may be inapplicable. Considering implementation potential, a system needs to be identified using only measured response information.

Initially, the team proposed an algorithm known as Generalized Iterative Least-Squares with Unknown Input (GILS-UI). To start the iteration process and

since excitation information is assumed to be unknown or not measured, it can be assumed to be zero. The acceleration response time-histories are recorded at high sampling rates, sometime at 4000 Hz giving a constant time increment of 0.00025 s. Initially, the team (Wang and Haldar, 1994) considered zero input excitation at first few time points. Later (Katkhuda et al., 2005) observed that the algorithm converged more efficiently if excitation time-histories were considered to be zero at all time points. Since the algorithm is ILS-based, its convergence capability can also be questioned. While conducting experimental verification of ILS-UI for a one dimensional beam, Vo and Haldar (2003) identified several sources of errors. They observed that the use of over-determined system of equations and post-processed and appropriate acceleration time-histories would assure convergence.

Integration of Measured Acceleration Time-Histories

It is not uncommon to hear that a SI-based concept will fail to identify a structure if measured acceleration time-histories are used, implying that it is inapplicable for the HAM purpose (Maybeck, 1979). The team observed that the post-processing of measured acceleration time-histories is essential to implement the SI-based concept. A typical accelerometer's raw signal output is shown in Fig. 1.
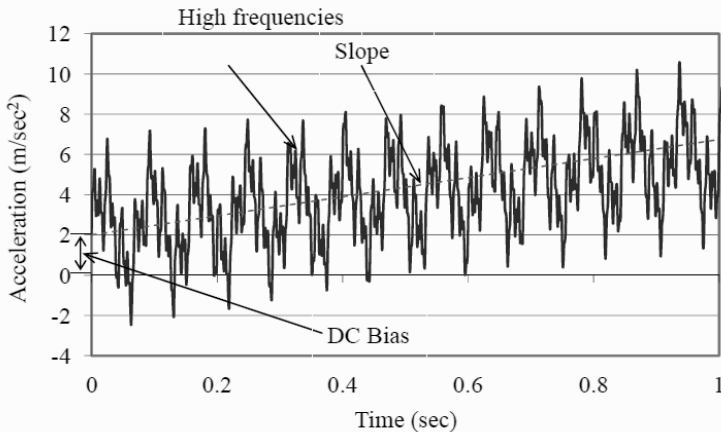
Fig. 1 Typical accelerometer's raw signal (Martinez-Flores and Haldar, 2005).

It may contain many sources of errors, including noise, high-frequency content, slope, and DC bias. All of them may not be present in all recorded acceleration time-histories. Noise is always present in the raw data even when measured by smart sensors. A low-pass filter is usually used to remove noise. Unwanted high-frequency responses can also be present in the raw data. Sometimes, high-frequency response is misclassified as noise. The Fast Fourier Transform (FFT) can be used to detect the presence of high-frequency noise. Filtering techniques can be used to remove them. Low-pass filters are used to remove high-frequencies and

high-pass filters are used to remove low-frequencies. In reality, the idea of low-pass filter is unobtainable. A type of filter known as a high-order Butterworth filter (Geis, 1989) provides the closest approximation. This is discussed in more detail by Vo and Haldar (2003).

Health Assessment with Post-Processed Data

The team observed that post-processed data failed to identify structures (Vo and Haldar, 2004). The algorithm failed to converge and gave negative stiffness for some of the elements. The amplitude error (scale factor and cross-coupling) and phase-shift error, shown in Fig. 2, were found to be responsible. Each accelerometer comes with a calibration factor to convert accelerometer's output voltage into acceleration unit. The provided scale factor always has calibration error embedded in it causing amplitude error. Another source of error that directly contributes to the amplitude error is the misalignment of the sensing element mounted inside the accelerometer's case. This error is often referred to as cross-coupling. For the accelerometers used by the research team, the combined root-sum-squared error for scale factor error and cross-coupling is ±2.8%. The algorithm will converge up to a convergence threshold of 0.5%. Thus, the post-processed acceleration time-histories will fail to identify a structure.

| (a) Amplitude error | (b) Phase-shift error |
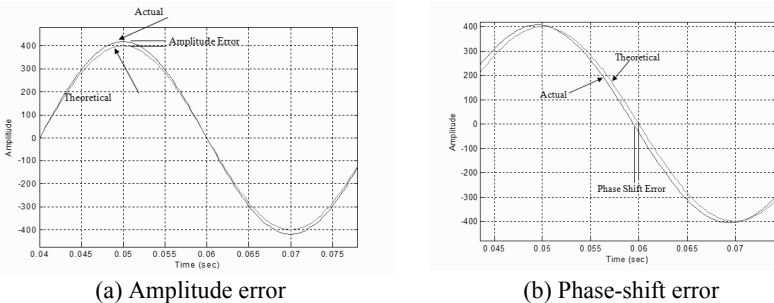
Fig 2. Primary sources of errors in a signal (Vo and Haldar, 2003).

The phase-shift error, as shown in Fig. 2(b), is generally caused by data latency and the numerical integration of noisy signal. However, the primary source of the phase-shift error is the integration of the measured acceleration time-histories to obtain velocity and displacement time-histories. The worst case phase-shift for velocity was estimated to be 1.8 degrees. Displacement response phase-shift was estimated to be a maximum of 6.5 degrees. The second source of phase-shift error is data latency caused by the sampling rate of the data-logger. This error occurs because there is a time delay in the sampling of two consecutive responses. The data-logger used in the test has a maximum latency of one micro-second. For 5 channels of data recording, a total latency of 5 micro-seconds is expected. For the tests, the phase shift error of data latency is found to be 0.09 degrees and integration error of 6.5 degrees. The algorithm converges only up to 0.5 degrees (Vo and Haldar, 2003).

In summary, the algorithms are found to be sensitive to both amplitude and phase-shift errors. The research team observed that by using fewer nodal responses, the total amplitude error can be reduced. The phase-shift error can be mitigated by scaling responses of all nodes based on the responses of a single reference node. The reference node can be chosen arbitrarily; however, the team used the excitation node at the reference node. Vo and Haldar (2004) discussed related issues in more details. Martinez-Flores (2005) conducted experimental verification of a scaled two-dimensional frame as shown in Fig. 3. In order to mitigate the two prime sources of errors, the author measured translational acceleration time-histories at all six node points. Then, the angular response time-histories were generated based on the proportionality of the transverse to angular responses. The proportionality constant of each rotational DDOF in the frame was estimated with respect to reference nodal response at node 2. Theoretically generated and experimental measured angular responses using an autocollimator at node 1 are shown in Fig. 4, indicating the validity of the response scaling approach.

Numerical Integration of Acceleration Time-Histories

The structural health assessment requires that the acceleration and corresponding velocity and displacement time-histories must be available. Since, only acceleration time-histories are measured (at high sampling rates), they need to be successively integrated to obtain the velocity and displacement time-histories. In general, numerical integration of discrete time-domain data involves two steps:
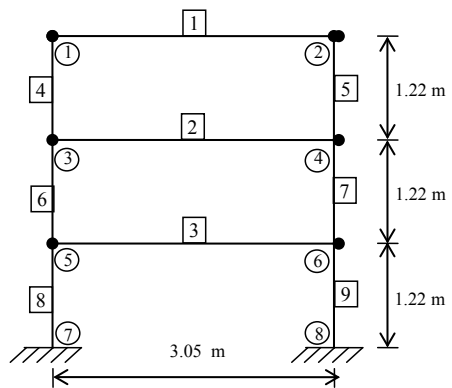
(a) pictorial view   (b) FE representation

Fig. 3 Experimental verification of a 2D frame (Martinez-Flores and Haldar, 2005).

Step 1 – Approximating a curve passing through successive data points with either a global interpolating polynomial defined over the whole domain of integration or a

collection of local or piecewise interpolating polynomials defined over the intervals that are divided by a small group of data points.

Step 2 – Integration of the interpolating polynomials over their defined individual domain.
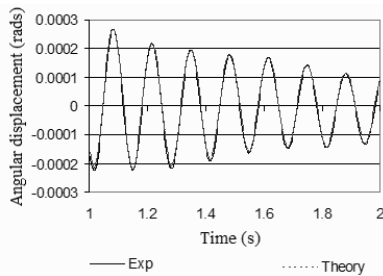
Fig. 4 Comparison between experimental and theoretical angular response at node 1 (Martinez-Flores and Haldar, 2005)

Several relatively simple and widely used numerical integration methods are available, including the trapezoidal rule, Simpson's rule, and Boole's rule (Davis and Rabinowitz, 1984). The mid-point rule and other open Newton-Cotes rules may not be appropriate since they do not consider end points. Vo and Haldar (2003) observed that it does not matter which method of numerical integration is used as long as the accelerometer data is sampled at a uniform interval of less than 0.01 s; it is expected to be at least two orders of magnitude less than that of the accelerometer errors.

System Identification of Large System

The use of GILS-UI implies that response information must be available at all DDOFs. It is practically impossible and uneconomical to install sensors at all DDOFs of a large realistic system. Considering accessibility and other practical issues, only small part(s) of a large system can be instrumented. This will lead to the introduction of the substructure(s) approach. Since response information is available at all DDOFs of the substructure, it can be identified using GILS-UI. However, it is not widely known that the convergence criteria used is based on the identification of the unknown excitation time-history at all time points with a pre-determined tolerance level. Thus, the use of GILS-UI also permits to identify the unknown excitation time-history. The implication of this observation will be discussed later.

The issues related to the sizes and numbers of substructures to be used to identify the health of a large system are open questions. For economic reason, they should be kept to an absolute minimum. They are discussed in detail by Das and Haldar (2012a, b). They observed that the algorithm may fail to identify a large system with minimum amount of information. After an inspection using the substructure approach and if the measured responses fail to identify the whole system, some additional steps are necessary to implement the concept. They observed that additional responses can be generated from the measured responses by a scaling approach (Vo and Haldar, 2004; Das and Haldar, 2012a, b). In widely used Extended Kalman filter (EKF)-based algorithms, the stiffness parameters for all the elements in

a system can be predicted using responses measured at limited DDOFs. However, the absolute minimum numbers of required responses are not known or discussed in the literature. When the number of responses is insufficient, additional responses can be generated by scaling. Das and Haldar (2012b) showed that for a 3-D frame represented by 72 DDOFs and with a substructure of 24 DDOFs, the maximum error in identification improves significantly when 3 additional scaled responses are used.

Most of the algorithms identify defects more efficiently if a substructure is close to them. Since, in general, the locations of defects are not known before an inspection, it will be desirable to use several substructures at different locations of a large structure. This will assure that at least one of the substructures is close to a defect location. Das et al. (2012c) considered a three-bay three-story 2-D frame. They used two small substructures: one at the top story and other at the first story level. Single and multiple defects were introduced at different locations relative to the substructures. The defective members were identified accurately in all cases.

Use of Extended Kalman Filter (EKF) Approach to Identify a Large Mildly Nonlinear System

The use of EKF concept in the health assessment of large structural systems using dynamic response information satisfying applicable governing differential equation also needs additional discussion. For proper implementation of the EKF concept, the excitation force vector and the initial state vector must be known. Thus, the use of basic EKF concept is inapplicable for the health assessment of large structures. The team observed that if EKF concept is integrated with the GILS-UI concept, both deficiencies can be removed very efficiently. It will require a two-stage approach. In stage 1, based on the measured response information, a substructure can be defined. Then, using the GILS-UI algorithm, the stiffness parameters of all the elements in the substructure, the time-history of the unknown excitation and damping information can be extracted. Suppose a substructure contains of one beam and another column element. In most structures, the section properties of beams and columns are expected to be similar. Thus, the identified properties of the beam and column in the substructure can be assigned to all other beams and columns appropriately. The identified damping can be used for the whole structure.

Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF)

The EKF and UKF algorithms can be used for HAM in the presence of nonlinearity. However, there are a number of challenges. The EKF procedure suffers several drawbacks in the linearization process. First, derivation of Jacobian matrices can introduce high implementation difficulties. Second, the filter can be unstable if sampling times are not sufficiently small. Third, a very small sampling time may increase computation time. The UKF procedure overcomes these issues by deterministically generating a set of sampling (sigma) points around the current state estimate. Then, these points are explicitly propagated through the nonlinear system equations to obtain more accurate estimation of the mean and covariance of the mapping results. The basic difference between EKF and UKF methods is in the prediction phase. It will be discussed during the presentation.

CONCLUSIONS

It is widely reported that large structural systems cannot be identified using measured response information. However, the authors used the concept in developing several HAM procedures. They used several algorithmic and computational tools to mitigate the issues. Some of them are discussed in this paper. The authors showed that SI-based procedures can be used for HAM if appropriate steps are taken.

REFERENCES

Das, A. K., and Haldar, A. (2012a). "Health assessment of three dimensional large structural systems – A novel approach." *Life Cycle Reliability and Safety Engineering*, 1(1), 1-14.

Das, A. K., and Haldar, A. (2012b). "Health assessment with only limited noise-contaminated response information." *Health Assessment of Engineered Structures: Bridges, Buildings and Other Infrastructures*, Edited by A. Haldar, World Scientific Publishing Co., Inc., New York.

Das, A. K., Haldar, A., and Chakraborty, S. (2012c). "Health assessment of large two dimensional structures using minimum information – recent advances." *Advances in Civil Engrg.*, 2012, Article ID 582472, doi:10.1155/2012/582472.

Davis, P., and Rabinowitz, P. (1984). *Method of numerical integration, 2nd Ed.,* Academic Press.

Geis, L. J. (1989). *Transform Analysis*, Prentice Hall, Englewood Cliffs, N. J.

Katkhuda, H., Martinez-Flores, R., and Haldar, A. (2005). "Health assessment at local level with unknown input excitation," *J. Struct. Engineering, ASCE,* 131(6), 956-965.

Katkhuda, H., and Haldar, A. (2008). "A novel health assessment technique with minimum information." *Struct. Cont. & Health Monitoring*, 15(6), 821-838.

Ling, X., and Haldar, A. (2004). "Element level system identification with unknown input with Rayleigh damping." *J. Engrg. Mech.*, *ASCE*, 130(8), 877-885.

Martinez-Flores, R. and Haldar, A. (2005). "Damage assessment potential of a novel system identification technique - experimental verification." Report No. CEEM-05-001, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona, 225 pages (based on Dr. Martinez's Ph.D. dissertation).

Maybeck, P. S. (1979), *Stochastic models, estimation, and control theory*, Academic Press, Inc., UK.

Vo, P. H., and Haldar, A. (2003). "Post-Processing of linear accelerometer data in structural identification." *J. Struct. Engineering,* 30(2), 123-130.

Vo, P., and Haldar, A. (2004). "Health assessment of beams - theoretical and experimental investigation." *J. Struct. Engrg.*, 31(1), 23-30.

Wang, D., and Haldar, A. (1994). "An element level SI with unknown input information." *J. Engrg. Mech., ASCE,* 120(1), 159-176.

Wang, D., and Haldar, A. (1997). "System identification with limited observations and without input." *J. Engrg. Mech.,, ASCE,* 123(5), 504-511.

Computational Modeling of Glass Panels for Mitigating the Injuries due to Air Blast

Hossein Ataei[1], James C. Anderson[2]

[1] Professional Engineer and PhD Candidate of Civil Engineering; ataei@usc.edu
[2] Professor of Civil Engineering; jamesa@usc.edu
Sonny Astani Department of Civil and Environmental Engineering
Viterbi School of Engineering, University of Southern California (USC)
3620 S. Vermont Ave., Kaprielian Hall Room 210, Los Angeles, CA, 90089

ABSTRACT

Glass fragments are a prime source of injury to occupants of buildings subjected to explosive events during which window glass breaks into flying shards or fragments that account for injuries ranging from minor cuts to severe wounds. A successful blast-resistant glazing design considers the principles of a reasonable degree of protection against explosive threats based on the proposed level of security and previous lessons learned. It also requires balancing of the safety and security of the window panels with physical appearance and cost-effectiveness. Therefore, it is necessary to have a better understanding of the response of glass panel to different glazing system parameters and blast load characteristics. In this study, an explicit finite element analysis is used in conjunction with fracture micromechanics' principles to model the failure of annealed and fully tempered glass panels. The response of the conventionally-framed window panels to blast loading cases is dependent upon the glazing material, blast intensities and glass imperfection locations and types. These results will help developing a more comprehensive flying glass injury model that will eventually enable the decision makers to use the appropriate structural, architectural and building perimeter choices to better address the threats facing the safety of the occupants during an air blast.

INTRODUCTION

Extreme loading on various glazing systems, during an air blast, results in shattering of glass panels into smaller pieces which move at very high velocities. Strict criteria should be in place for design of building envelope systems and structural glazing to resist blast waves because of their high-impact low-frequency nature and potential devastating toll on the occupants. To study the effects of glass non-homogeneity on window panel fracture patterns, the blast pulse is assumed to have a triangular shape: The overpressure rises instantaneously to its peak value and has a linear decay back to zero (atmospheric pressure) in a very short period of time.

In this paper, conventionally-framed annealed glass (AG) and fully tempered (FT) window panels of 50 cm x 75 cm (20" x 30") are modeled and analyzed. Each window panel is studied for the effects of two different air blast loadings: Smaller

intensity air blast due to 100 Lbs of TNT and a Higher intensity blast due to 1000 Lbs of TNT. For both explosive weights above, the explosion stand-off distance is set to be 30.5 meters (100 ft). This distance, according to the Defense Threat Reduction Agency, will cause severe wounds and casualties by flying glass shards. The blast load characteristics: Peak pressure (P_r), Reflected impulse (i_r) and the Load duration (t_0) are summarized in Table 1, below, using the software "A.T. Blast" and "Defense Threat Reduction Agency" manuals given the Weight of explosive (*W*) and Stand-off distances (*R*) for the case where angle of attack (α) equals zero.

Table 1. **Blast Loading Characteristics [Ataei H., Anderson J.C., (2012)].**

W (Kg)	R (m)	i_r (Pa . msec)	P_r (Pa)	t_0 (msec)
45.0	30.5	249636.85	39782.77	12.55
453.5	30.5	1249421.52	165267.40	15.12

MODEL MATERIAL PROPERTIES

Annealed Glass (AG) is the most common type of glass used in conventional construction. Subjected to an air blast, it fractures in irregular patterns leading to sharp glass shards that cause injuries to the building occupants varying from minor cuts to severe wounds [GSA Standard Test Methods]. Fully Tempered (FT) glass is made by controlling uniform heating and rapid cooling of the annealed glass. FT glass has more hardening and about four times more strength of the annealed glass and it has a different pattern of shattering which tends to dice it into smaller fragment pieces as opposed to the annealed glass shards [Johnson N.F., (2006)].

The window panel, in this paper, is modeled using two different types of glazing materials: (1) Annealed Glass (AG) with 6.0 mm (1/4") thickness and (2) Fully Tempered (FT) glass window panels with the thickness of 7.5 mm (0.3"). The types and the properties of the model materials are summarized in Table 2.

Table 2. **Glazing Material Properties.**

Glazing Material	Elastic Modulus (GPa)	Poisson's Ratio	Density (kg/m^3)	Fracture Strength (MPa)
Annealed Glass	69	0.22	2491.19	27.58
Fully Tempered Glass	69	0.22	2491.19	110.32

Failure stresses of glass can vary among the glass samples even from the same lot. The damage accumulation factor, glass age and the shape and location of non-homogeneity and imperfections can negatively affect the actual failure strength of glass panels. Therefore, non-homogeneity of structural glass panels and the glass properties variability need to be further investigated in development of flying glass injury models. These models will better address the crack propagation patters across the window panels to evaluate more comprehensively the potential hazard level of each case on the building occupants' safety [Ataei H., Anderson J.C., (2012)].

EXPLICIT FINITE ELEMENT ANALYSIS STABILITY

Air blasts, as high-speed short-duration dynamic loads, are usually solved using explicit methods that do not require global tangent stiffness matrix [Belytscheko et al., (2005)]. Explicit dynamic method requires a stability limit which technically is a small time increment that is independent of the type and duration of the loading [Ataei et al., (2007)] and solely depends on the highest natural frequency of the model (ω_{max}) and the fraction of critical damping (ξ_{max}) in the mode with the highest frequency as per below equation [ABAQUS User's Manual, (2011)]:

$$\Delta t_{stability} \leq \frac{2}{\omega_{max}} \left(\sqrt{1 + \xi_{max}{}^2} - \xi_{max} \right)$$

A more conservative and practical method to calculate the stability limit of the explicit analysis is based on the smallest element length in the mesh (L_{min}) and the dilatational wave speed of the material (c_d) [ABAQUS User's Manual, (2011)]:

$$\Delta t_{stability} \approx \frac{L_{min}}{c_d}$$

As the material yields, the wave speed reduces and therefore, the stability limit can take larger values after yielding which will result in faster convergence. [Belytscheko et al., (2005)]. Larger element sizes will increase the stability limit, however, finer mesh sizes are often necessary to obtain more accurate analysis results [Belytscheko et al., (2005)]. The best approach is to have a fine mesh that is as uniform as possible to obtain the highest possible stability limit [Ataei H, Anderson J.C., (2012)]. In this paper, in order to study the crack propagation patterns, a very fine mesh that consists of square 4-noded shell elements is employed to model the entire window panel as a 3-dimensional deformable shell.

YIELD LINE ANALYSIS OF RECTANGULAR WINDOW PANELS

Figure 1 (below) demonstrates the fracture pattern of rectangular window panels with (a x b) dimensions at which a is the shorter dimension.
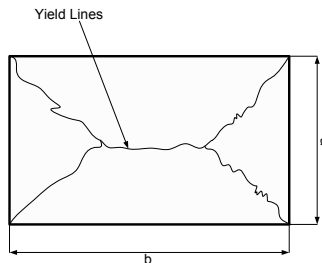
Figure 1. Theoretical yield-line pattern of rectangular window panels at failure.

For a rectangular window panel that is subjected to uniform blast pressure of (w) and is simply-supported (conventional framing) along its four edges, the maximum moment that the panel can take is [Nawy E.G., (2000)]:

$$M = \frac{wa^2}{24}\left[\sqrt{3+\left(\frac{a}{b}\right)^2} - \frac{a}{b}\right]^2$$

Therefore, given the ultimate strength of the window panels, the breakage pressure values for each window that make the cracks start and propagate afterwards can be calculated [Reddy J.N., (2006)]. As summarized in Table 3 (below), based on the yield line analysis of simply supported glazing systems, the amount of blast overpressure that makes each of our models fail can be calculated. Moreover, it can be determined whether the breakage does happen in the window panels. However, the yield-line analysis only estimates the ultimate failure load on the panel and does not provide insight on the exact time at which the crack initiates nor on its propagation patterns across the panel versus time due to highly dynamic nature of crack propagation problems [Ataei H., Anderson J.C., (2012)].

Table 3. Yield Line Analysis of Conventional 50cm x 75cm Window Panels.

		Annealed Glass (50 cm x 75 cm)	Fully-Tempered Glass (50 cm x 75 cm)
45 Kg TNT at 30.5 m	Max Pressure (Pa)	39782.77	39782.77
	Failure Pressure (Pa)	7550.00	47301.35
	Breakage Happens	*Yes*	*No*
453.5 Kg TNT at 30.5 m	Max Pressure (Pa)	165267.40	165267.40
	Failure Pressure (Pa)	7550.00	47301.35
	Breakage Happens	*Yes*	*Yes*

FINITE ELEMENT ANALYSIS OF GLASS FRACTURE DYNAMICS

In order to precisely capture the behavior of window panels that are subjected to air blast, an infinitesimal time increment of 0.0125 msec (0.0000125 sec) has been adopted to advance the state of the model in the explicit analysis. During the analysis, the elements that reach local failure (stress; strain or crack displacement) are removed from the mesh before the whole model advances to the next timeframe. Henceforth, the dynamic element deletion sequence is visualized as the crack propagates.

The fracture patterns of annealed and fully-tempered glass panels are modeled with brittle failure option in ABAQUS. The maximum stress criterion is utilized to identify the glass failure. This criterion states that failure occurs when the maximum principle stresses in two dimensional coordinate system, (σ_1, σ_2), reach either the maximum uniaxial tension (σ_t) or compression strength (σ_c) of the glass.

The American standards (ASTM E2461) and Canadian standards (CAN/CGSB-12.20-M89) limit the allowable maximum fracture stress for annealed glass to 25 MPa in the center of the pane and 20 MPa on the edges. For fully tempered glass, however, both standards suggest the allowable stress be no greater than 100 MPa in the window panel center. More accurate guidelines limit the glass deflection to the least of either span length divided by 175 or 3/4" (whichever be

less). Moreover, the shock tube experiments performed by Kumar and Shukla (2011) propose an allowable strain of 0.01% for annealed glass and 2% for FT panels before fracture.

HOMOGENEOUS WINDOW PANEL FAILURE PATTERNS

The calculated failure pattern of conventional homogeneous 50 cm x 75 cm annealed glass window panels, subjected to small intensity air blast, is illustrated in Figure 2. The stress contours in the corners rise at 2.5 msec and the stress waves find their way towards the panel center at 4.175 msec. Then, at 4.75 msec, the glass panel cracks in the middle along the vertical yield line and the crack propagates towards the upper and lower boundaries. As predicted by the yield-line analysis, the cracks expand towards the corners at 6.75 msec followed by the total failure at 12.00 msec.
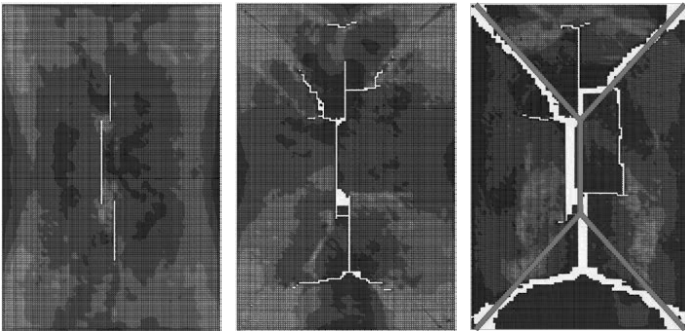
Figure 2. Stress distribution and crack propagation patterns for conventional homogeneous AG panels subjected to small intensity air blast: (left) 4.75 msec, (center) 6.75 msec, (right) 12.00 msec.

Failure pattern of conventional homogeneous annealed glass window panels subjected to higher intensity air blast is illustrated in Figure 3. The stress contours
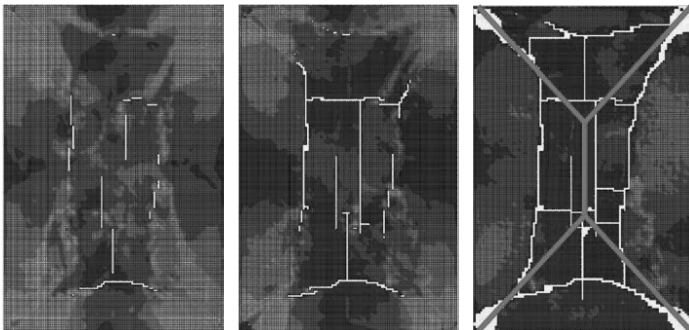
Figure 3. Stress distribution and crack propagation patterns for conventional homogeneous AG panels subjected to higher intensity air blast: (left) 3.75 msec, (center) 4.25 msec, (right) 5.85 msec.

across the panel rise with faster pace, compared to the case of smaller blast, and the stress waves find their way towards the panel center where a higher crack accumulation will be formed. The glass panel cracks in the middle at 3.75 msec and the crack propagates towards the corners at 4.25 msec followed by total failure that happens at 5.85 msec in compliance with the yield line analysis predicted pattern.

Figure 4 (below) demonstrates the failure pattern of conventional simply-supported homogeneous FT glass panels subjected to higher intensity air blast. The longitudinal vertical crack at 8.875 msec will start branching further at 11.363 msec forming other horizontal and vertical cracks and to eventually divide the panel into smaller blocks and fragments as the dominant failure pattern at 12.975 msec.
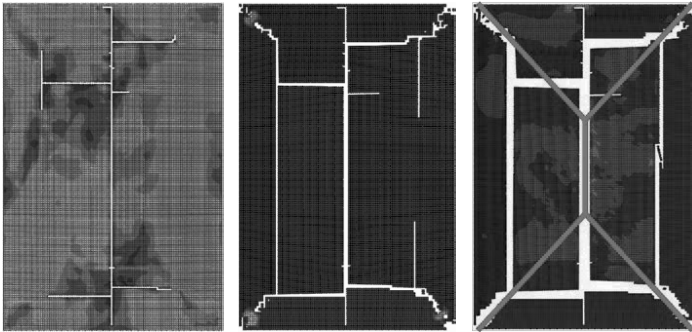
Figure 4. Stress distribution and crack propagation patterns for conventional homogeneous FT panels subjected to higher intensity air blast: (left) 8.875 msec, (center) 11.363 msec, (right) 12.975 msec.

EFFECT OF WINDOW PANEL DEFECTS ON FAILURE PATTERNS

ASTM C1036 (2011) categorizes major glass defects as crush; deep short scratch (dig); inclusion of small foreign matter particle in the surface of flat glass (dirt); entrapped air bubbles; blemishes; scratches; as well as shell chips (circular indentation in the glass edge); and v-chips that are v-shaped imperfections in the edge of the glass lite. Annealed glass panels with a thickness of 6 mm (1⁄4") or less shall not vary in thickness more than 0.1 mm (0.004") over a 100 mm (4") length.

Imperfections of fully-tempered glass panels, according to ASTM C1048 (2012), are commonly: curvature warps across the entire one or both panel dimensions; crush; deep short scratches (dig); small particle of foreign matter (sand or other imperfections on the roller surface during the heating or cooling procedure); distortion due to heating; air bubble inclusions; scratches and holes. The FT panel holes need to have a minimum distance of greater value of 6 mm (1/4") or 2 times of panel thickness from any panel edge as well as greater of 10 mm (3/8") or 2 times of thickness from the adjacent hole rim. Minimum diameter of the holes should be greater of either 6 mm (1/4") or the panel thickness.

In this paper, we have studied and compared the effects of two defect categories on failure patterns of the proposed simply-supported 50 cm x 75 cm

window: (1) Two chips in the right-hand-side and upper-left edges of the annealed glass subjected to smaller intensity air blasts; (2) existence of three holes on fully-tempered glass panel surface (in compliance with ASTM C1048 five-point examination locations) subjected to higher intensity blast loads.

Figure 5 (below) illustrates the crack propagation path for defective conventional annealed glass panels that are subjected to small intensity air blast. As it is seen in the figure, the window panel center starts cracking at 4.362 msec and the crack propagates vertically at 5.325 msec with maximum stress lines passing through the corners and the upper-left chip. Due to the weakness of the panel in the right side, the bottom corner will come off at 7.050 msec (while the crack path is in general agreement with the yield line analysis). The panel will then fail at 8.725 msec which occurs faster compared to the homogeneous AG panel failure at 12.00 msec.
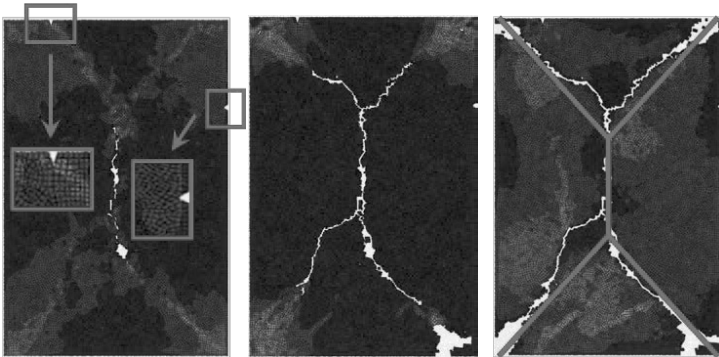
Figure 5. Stress distribution and crack propagation patterns for conventional defective AG panels subjected to small intensity air blast: (left) 5.325 msec, (center) 7.050 msec, (right) 8.725 msec.
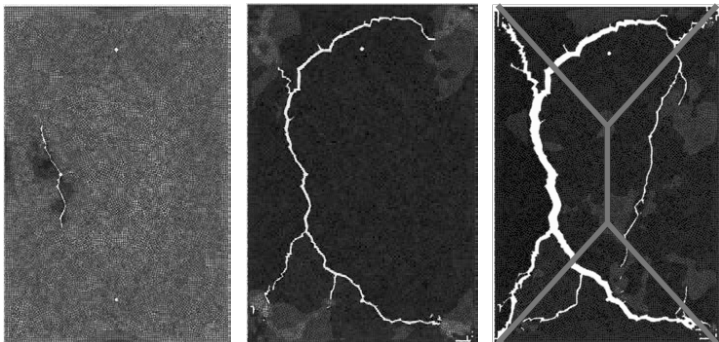
Figure 6. Stress distribution and crack propagation patterns for conventional defective FT panels subjected to higher intensity air blast: (left) 4.60 msec, (center) 7.787 msec, (right) 9.90 msec.

Figure 6 (above) demonstrates the crack propagation path and failure pattern of defective fully tempered glass panel with holes. The crack initiates from mid-panel hole location at 4.57 msec and propagates vertically at 4.60 msec. The crack path passes through the second hole (imperfection) as a weak point at 4.70 msec. Thereafter, cracks find their way towards the panel corners at 7.787 msec (as seen in the figure) until the complete panel failure at 9.90 msec which happens faster in comparison with homogeneous fully tempered glass panel failure at 12.975 msec.

SUMMARY AND CONCLUSION

The required blast overpressure amount for window breakage is calculated using the yield line analysis as well as predicting whether the failure will occur in each case. However, in order to envisage the crack paths and obtain the panel failure time histories, explicit finite element simulation is adopted for modeling the window failure. Following the crack path through removal of the failed elements, the real behavior of the material is therefore visualized at each time increment. Conventional simply-supported window panels with both annealed and fully tempered glazing technologies subjected to different blast loading cases are studied and the yield-line analysis results are compared with those of fracture mechanics principles using finite element simulations for both homogeneous and defective panels.

REFERENCES

ABAQUS Analysis User's Manual (2011). Ver. 6.11, ABAQUS Inc., RI, USA

ASTM C1036 - 11 (2011). "Standard Specification for Flat Glass"

ASTM C1048 - 12 (2012). "Standard Specification for Heat-Strengthened and Fully Tempered Flat Glass"

ASTM E2461 - 05 (2011). "Standard Practice for Determining the Thickness of Glass in Airport Traffic Control Tower Cabs"

Ataei H., and Anderson J.C. (2012) "Mitigating the Injuries from Flying Glass due to Air Blast", *ASCE 6th Congress on Forensic Engineering*, San Francisco, CA

Ataei H., and Anderson J.C. and Niazy A.M. (2007) "Resistance of Glass Panels to Air Blast", *PROTECT2007 International Congress*, Vancouver, BC, Canada

A.T-Blast 2.2, *Applied Research Associates Inc.*, Albuquerque, NM, USA

Belytscheko T., and Liu W.K., and Moran B. (2005) "Nonlinear Finite Element for Continua and Structures", John Wiley and Sons Ltd.

Canadian General Standards Board (1989). "Structural design of glass for buildings", CAN/CGSB-12.20-M89

Johnson N.F. (2006). "International Standards for Blast Resistant Glazing", *Journal of ASTM International*, Vol. 3, No. 4

Nawy E. G. (2000). "Reinforced Concrete; A Fundamental Approach", 4th Edition, 2000, Prentice Hall, Upper Saddle River, NJ

Reddy J. N. (2006). "Theory and Analysis of Elastic Plates and Shells", 2nd Edition, 2006, CRC Press

United States General Services Administration (2003). "Standard Test Method for Glazing and Window Systems Subject to Dynamic Overpressure Loadings"

Computationally Efficient Control Design for a Smart Structure with Uncertainty

M. Kamalzare[1], E.A. Johnson[1], M. ASCE, and S.F. Wojtkiewicz[2], M. ASCE

[1]Sonny Astani Department of Civil and Environmental Engineering,
University of Southern California, 3620 S Vermont Ave, Los Angeles, CA 90089
[2]Department of Civil Engineering, University of Minnesota,
500 Pillsbury Drive S.E., Minneapolis, MN 55455
email: kamalzar@usc.edu, JohnsonE@usc.edu, ykvich@umn.edu

ABSTRACT

Optimal structural control design for controllable passive dampers is difficult and computationally expensive. Even with a linear structure, the controlled system is nonlinear since the control appears parametrically and the damper that can only exert dissipative forces. The authors previously introduced a novel method, based on a low-order nonlinear Volterra integral equation (NVIE), that provides a computationally efficient means to simulate responses for such systems. Thus far, the authors have considered deterministic models with semiactive control devices; however, uncertainties in key structural elements cannot be neglected in real-world systems and may intensify the computational burden, generally demanding a Monte Carlo simulation to characterize the effects of uncertainty. Herein, a 100 degree-of-freedom isolated building model with a controllable damper in the isolation is studied with uncertainty in the isolation rubber bearing stiffness and damping. Numerical results confirm the accuracy, stability, and computational efficiency of the proposed simulation methodology.

INTRODUCTION

Controllable passive dampers, such as variable orifice or magnetorheological fluid dampers (Housner et al., 1997; Symans and Constantinou, 1999; Soong and Spencer, 2002), have been investigated in recent years due to: (1) their ability to mitigate the structural responses dramatically compared to the passive controllers since they can utilize global system response information; (2) the assurance of a stable system since they cannot apply non-dissipative forces; and (3) their minimal external energy dependence. Due to the typically parametric nature of most control laws for such devices, many full nonlinear simulations of system response with different control strategies are often required. The result is a rather inefficient and time-consuming design procedure, particularly for structures of real-world complexity with high-order models.

The authors introduced a method (Gaurav et al., 2011), based on a nonlinear Volterra integral equation in non-standard form, for computationally efficient analysis of large, complex systems that are linear and deterministic except a few local nonlinear or uncertain elements. Further, the authors demonstrated (Kamalzare et al., 2013) how the equation of motion of a structure with a semiactive control device can be written in a form such that the proposed approach can be implemented to efficiently design optimal semiactive control strategies. For a 100 degree-of-freedom (DOF) structure model with clipped-optimal control of an ideal controllable damper, the computational cost of a control design parameter study was reduced by more than two order of magnitudes. However, this controlled system was entirely deterministic.

Herein, uncertainty is introduced into the system model, using a Monte Carlo simulation to characterize the effects of uncertain elements. The design of a semiactive control strategy is studied for the same 100 DOF building model, including a superstructure on a base isolation layer. The stochastic behavior is induced by uncertain isolation layer structural parameters (the rubber bearing stiffness and damping); the control force is provided by a controllable damper in the isolation layer. Numerical results confirm the accuracy, stability and computational efficiency of the proposed algorithm.

METHODOLOGY

The authors introduced (Gaurav et al., 2011) an approach for efficient analysis of linear systems with a few local modifications. This approach evaluates the effect of the local modifications (by solving a low-order, generally nonlinear, Volterra integral equation) and replaces them with the equivalent force. The resulting equation of motion is linear and can be solved easily using superposition. This section adapts this approach for efficient uncertainty analysis in the systems controlled by a semiactive control device.

Consider a nominal linear deterministic structure model in state space form

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{w}(t), \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{1}$$

where $\mathbf{x}(t)$ is the $n_\mathbf{x} \times 1$ state vector; $\dot{(\ )}$ denotes a derivative with respect to time; $\mathbf{A}$ is the $n_\mathbf{x} \times n_\mathbf{x}$ system matrix; $\mathbf{B}$ is the $n_\mathbf{x} \times n_\mathbf{w}$ excitation distribution matrix; $\mathbf{w}(t)$ is the $n_\mathbf{w} \times 1$ external excitation vector; and $\mathbf{x}_0$ is the initial condition.

Let $\mathbf{u} = \mathbf{g}^\mathbf{u}(\mathbf{X})$ be a $n_\mathbf{u} \times 1$ vector of control device forces that will be used to mitigate the structural responses. Additionally, let $\mathbf{f} = \mathbf{g}^\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ be a $n_\mathbf{f} \times 1$ vector of forces within the structure caused by the uncertainty in parameter vector $\boldsymbol{\theta}$. Incorporting these forces, the equations of motion can be written as follows

$$\dot{\mathbf{X}}(t) = \mathbf{A}\mathbf{X}(t) + \mathbf{B}\mathbf{w}(t) + \mathbf{L}^\mathbf{u}\mathbf{g}^\mathbf{u}(\mathbf{X}(t)) + \mathbf{L}^\mathbf{f}\mathbf{g}^\mathbf{f}(\mathbf{X}(t), \boldsymbol{\theta}), \qquad \mathbf{X}(0) = \mathbf{x}_0 \tag{2}$$

where $\mathbf{X}(t)$ is the $n_\mathbf{x} \times 1$ state vector; $\mathbf{L}^\mathbf{u}$ is an $n_\mathbf{x} \times n_\mathbf{u}$ constant influence matrix that maps the nonlinear control force to the appropriate system states; $\mathbf{L}^\mathbf{f}$ is an $n_\mathbf{x} \times n_\mathbf{f}$ constant influence matrix that maps the uncertainty effect to the appropriate system states; and the initial condition remains the same as that of the nominal system.

The functional forms of $\mathbf{g}^\mathbf{u}(\mathbf{X})$ and $\mathbf{g}^\mathbf{f}(\mathbf{X}, \boldsymbol{\theta})$ depend on the nature of modifications (*i.e.*, the model used for semiactive control force and the form of the equation describing the uncertainty behavior) in the system and are not required to be in any specific form. Since these effects are localized, they are direct functions of only some subset

or linear combination of states, $\overline{\mathbf{X}}^{\mathbf{u}}(t) = \mathbf{G}^{\mathbf{u}}\mathbf{X}(t)$ and $\overline{\mathbf{X}}^{\mathbf{f}}(t) = \mathbf{G}^{\mathbf{f}}\mathbf{X}(t)$, respectively, where $\mathbf{G}^{\mathbf{u}}$ is a $n_{\overline{\mathbf{x}},\mathbf{u}} \times n_{\mathbf{x}}$ matrix and $\mathbf{G}^{\mathbf{f}}$ is a $n_{\overline{\mathbf{x}},\mathbf{f}} \times n_{\mathbf{x}}$ matrix.

The two modifications — control force and force from uncertainty — can be combined by merging the state subsets $\overline{\mathbf{X}}^{\mathbf{u}}$ and $\overline{\mathbf{X}}^{\mathbf{f}}$ on which the control and uncertainty depend, respectively, and combining the modification influence matrices $\mathbf{L}^{\mathbf{u}}$ and $\mathbf{L}^{\mathbf{f}}$ of the forces from control and uncertainty, respectively. If the state subsets are independent, then one may define a corresponding combined $n_{\overline{\mathbf{x}}} \times 1$ vector $\overline{\mathbf{X}} = [(\overline{\mathbf{X}}^{\mathbf{u}})^{\mathrm{T}} \ (\overline{\mathbf{X}}^{\mathbf{f}})^{\mathrm{T}}]^{\mathrm{T}} = \mathbf{GX}$, with $n_{\overline{\mathbf{x}}} \times n_{\mathbf{x}}$ matrix $\mathbf{G} = [(\mathbf{G}^{\mathbf{u}})^{\mathrm{T}} \ (\mathbf{G}^{\mathbf{f}})^{\mathrm{T}}]^{\mathrm{T}}$, where $n_{\overline{\mathbf{x}}} = n_{\overline{\mathbf{x}},\mathbf{u}} + n_{\overline{\mathbf{x}},\mathbf{f}}$; if there are elements of $\overline{\mathbf{X}}^{\mathbf{u}}$ and $\overline{\mathbf{X}}^{\mathbf{f}}$ in common, then the redundant elements of $\overline{\mathbf{X}}$ and redundant rows of $\mathbf{G}$ may be removed, leaving $n_{\overline{\mathbf{x}}} < n_{\overline{\mathbf{x}},\mathbf{u}} + n_{\overline{\mathbf{x}},\mathbf{f}}$. Similarly, if the influence matrices $\mathbf{L}^{\mathbf{u}}$ and $\mathbf{L}^{\mathbf{f}}$ have unique columns, then they can be combined into the $n_{\mathbf{x}} \times n_{\mathbf{p}}$ influence matrix $\mathbf{L} = [\mathbf{L}^{\mathbf{u}} \ \mathbf{L}^{\mathbf{f}}]$, with the resulting combined $n_{\mathbf{p}} \times 1$ modification vector $\mathbf{g}(\overline{\mathbf{X}}) = [\{\mathbf{g}^{\mathbf{u}}(\overline{\mathbf{X}}^{\mathbf{u}})\}^{\mathrm{T}} \ \{\mathbf{g}^{\mathbf{f}}(\overline{\mathbf{X}}^{\mathbf{f}})\}^{\mathrm{T}}]^{\mathrm{T}}$, where $n_{\mathbf{p}} = n_{\mathbf{u}} + n_{\mathbf{f}}$; if columns of the influence matrices are in common, then redundant columns of $\mathbf{L}$ can be eliminated and the corresponding rows of $\mathbf{g}(\overline{\mathbf{X}})$ added together, resulting in $n_{\mathbf{p}} < n_{\mathbf{u}} + n_{\mathbf{f}}$.

Then, (2) can be written as

$$\dot{\mathbf{X}}(t) = \mathbf{AX}(t) + \mathbf{Bw}(t) + \mathbf{Lg}(\overline{\mathbf{X}}(t)), \qquad \mathbf{X}(0) = \mathbf{x}_0 \tag{3}$$

If the force $\mathbf{p}(t) = \mathbf{g}(\overline{\mathbf{X}}(t))$ were known *a priori*, (3) would be solved by superimposing the response to $\mathbf{p}(t)$ on the response to $\mathbf{w}(t)$, both for all states $\mathbf{X}$ and the subset $\overline{\mathbf{X}}$:

$$\mathbf{X}(t) = \mathbf{x}(t) + \int_0^t \mathbf{H_L}(t-\tau)\mathbf{p}(\tau)d\tau, \qquad \overline{\mathbf{X}}(t) = \overline{\mathbf{x}}(t) + \int_0^t \overline{\mathbf{H}}_{\mathbf{L}}(t-\tau)\mathbf{p}(\tau)d\tau \tag{4}$$

where $\mathbf{H_L}(t) = e^{\mathbf{A}t}\mathbf{L}$ is the response to an impulse in the pattern of modification $\mathbf{g}$, $\overline{\mathbf{x}}(t) = \mathbf{Gx}(t)$ is the nominal response subset, and $\overline{\mathbf{H}}_{\mathbf{L}}(t) = \mathbf{GH_L}(t)$ is the corresponding impulse response subset. Using (4), $\mathbf{p}(t) = \mathbf{g}(\overline{\mathbf{X}}(t))$ can be written as

$$\mathbf{p}(t) - \mathbf{g}\left(\overline{\mathbf{x}}(t) + \int_0^t \overline{\mathbf{H}}_{\mathbf{L}}(t-\tau)\mathbf{p}(\tau)d\tau\right) = \mathbf{0} \tag{5}$$

which is a $n_{\mathbf{p}}$-dimensional nonlinear Volterra integral equation (NVIE) in nonstandard form. The authors demonstrated (Gaurav et al., 2011) a discretization of the integral in (5), and a solution for $\mathbf{p}(t)$ and, subsequently, for the responses of the modified system. If $n_{\mathbf{p}}$ and $n_{\overline{\mathbf{x}}}$ are much smaller than $n_{\mathbf{x}}$, then solving (5) is much less computationally intensive than solving the full nonlinear system of ordinary differential equations.

SEMIACTIVE CONTROL FORCE

An "ideal" controllable damper can only exert a dissipative force that resists motion, *i.e.*, a force in the direction opposite from the velocity across the device. Assume some control design strategy (which is not the primary focus herein) provides a vector $\mathbf{u}^{\mathbf{d}}(t) = -\mathbf{K}^{\mathbf{d}}\mathbf{X}(t)$ of desired device forces, where $\mathbf{K}^{\mathbf{d}}$ is some state feedback control gain. The clipping algorithm to determine the force command for the ith device is

$$u_i = g_i^{\mathbf{u}}(\overline{\mathbf{X}}^{\mathbf{u}}) = u_i^{\mathrm{d}} H(-u_i^{\mathrm{d}} v_i) = \begin{cases} u_i^{\mathrm{d}}, & u_i^{\mathrm{d}} v_i < 0 \quad \text{(dissipate energy)} \\ 0, & u_i^{\mathrm{d}} v_i \geq 0 \quad \text{(cannot add energy)} \end{cases} \tag{6}$$

where $u_i^d(t)$ is the ith element of desired force vector $\mathbf{u}^d(t)$; $H(\cdot)$ is the Heaviside unit step function; and $v_i(t)$ is the velocity across the ith device. If the vector of device velocities is given by $\mathbf{v}(t) = \mathbf{VX}(t)$, then the control force function $\mathbf{g^u}(\cdot)$ is only dependent on $\overline{\mathbf{X}}^\mathbf{u} = \mathbf{G^u X}$ where $\mathbf{G^u} = [-(\mathbf{K}^d)^T \quad \mathbf{V}^T]^T$ is a $2n_\mathbf{u} \times n_\mathbf{x}$ matrix, and $g_i^\mathbf{u}(\overline{\mathbf{X}}^\mathbf{u}) = u_i^d H(-u_i^d v_i) = \overline{X}_i^\mathbf{u} H(-\overline{X}_i^\mathbf{u} \overline{X}_{i+n_\mathbf{u}}^\mathbf{u})$ for $i = 1, ..., n_\mathbf{u}$.

For a low order system with few controllable dampers, a parameter study over all elements of gain matrix $\mathbf{K}^d$ may be feasible. However, for any system of real size, such a parameter study would be a monumental task. An alternate approach is to study a family of control gains generated by parameterizing an optimal control design. Herein, a linear quadratic regulator (LQR) approach is used to compute a family of state feedback gains $\{\mathbf{K}_1^d, \mathbf{K}_2^d, ...\}$, each of which minimizes, for a linear actuator, the cost function

$$J_i = \int_0^T [\mathbf{Z}_i^T(t)\mathbf{Q}_i\mathbf{Z}_i(t) + \mathbf{u}_i^T(t)\mathbf{R}_i\mathbf{u}_i(t)]dt \tag{7}$$

where $\mathbf{Z}(t) = \mathbf{C_z X}(t) + \mathbf{D_z w}(t) + \mathbf{F_z p}(t)$ is a vector of responses to be regulated, with corresponding nominal responses $\mathbf{z}(t) = \mathbf{C_z x}(t) + \mathbf{D_z w}(t)$. For the optimization to be well-posed, the weighting matrices must satisfy $\mathbf{C_z^T Q_i C_z} \geq 0$ and $\mathbf{F_z^T Q_i F_z} + \mathbf{R}_i > 0$.

NUMERICAL EXAMPLE

This example uses the same 100-DOF model of a isolated building the authors used previously (Kamalzare et al., 2013) to demonstrate computationally efficient semi-active control design, except the isolation layer parameters are uncertain herein.

Model Description

The superstructure is an 11-story, 2-bay building modeled as a 99 DOF linear plane frame (horizontal, vertical and rotational degrees of freedom at each node, with three nodes per floor). The column weights are neglected as they are small relative to the floor weights. Consistent mass and stiffness matrices are used for the (floor and column) beam elements; Rayleigh damping is assumed for the superstructure, with 1st and 10th modal damping ratios both equal to 3%. The fundamental superstructure mode, if it were a fixed-base structure, has a 1.05 s period and 3% damping. The ground excitation $\ddot{x}_g$ is applied only in the horizontal direction. The base mass, linear isolator stiffness, and linear viscous isolator damping are chosen such that the isolation mode has a typical (Skinner et al., 1993) period $T_1^i = 2.76$ s and a 5.5% damping ratio. The building weight is $W = 1.28$ MN and height is $h = 44$ m. The isolation layer is constrained to move in the horizontal direction only. The resulting 100 DOF model has equation of motion

$$\begin{bmatrix} m_b & \mathbf{0}^T \\ \mathbf{0} & \mathbf{M_s} \end{bmatrix} \begin{Bmatrix} \ddot{x}_b \\ \ddot{\mathbf{x}}_s \end{Bmatrix} + \begin{bmatrix} c_b + \mathbf{r}^T\mathbf{C_s r} & -\mathbf{r}^T\mathbf{C_s} \\ -\mathbf{C_s r} & \mathbf{C_s} \end{bmatrix} \begin{Bmatrix} \dot{x}_b \\ \dot{\mathbf{x}}_s \end{Bmatrix} + \begin{bmatrix} k_b + \mathbf{r}^T\mathbf{K_s r} & -\mathbf{r}^T\mathbf{K_s} \\ -\mathbf{K_s r} & \mathbf{K_s} \end{bmatrix} \begin{Bmatrix} x_b \\ \mathbf{x}_s \end{Bmatrix}$$
$$= -\begin{bmatrix} m_b & \mathbf{0}^T \\ \mathbf{0} & \mathbf{M_s} \end{bmatrix} \begin{Bmatrix} 1 \\ \mathbf{r} \end{Bmatrix} \ddot{x}_g^a + \begin{Bmatrix} 1 \\ \mathbf{0} \end{Bmatrix} u \tag{8}$$

where x_b is the base displacement relative to the ground; m_b is the base mass; c_b and k_b are the isolation linear damping and stiffness, respectively; $\mathbf{x}_s$ is a vector of horizontal and vertical displacements and rotation of the superstructure nodes relative to the ground; $\mathbf{M_s}$, $\mathbf{C_s}$ and $\mathbf{K_s}$ are the superstructure mass, damping and stiffness matrices, respectively;

$\mathbf{r} = [1\,0\,0 \quad 1\,0\,0 \quad \cdots \quad 1\,0\,0]^{\mathrm{T}}$ (*i.e.*, ones in elements corresponding to horizontal displacements in $\mathbf{x}_s$) is the influence vector; and $u(t)$ is the control force. Ground excitation $\ddot{x}_g^a$ is the N-S El Centro record of the 18 May 1940 Imperial Valley earthquake sampled at 50 Hz.

Cost function (7) has control weight $R_i = W^{-2}$ and diagonal response weight $\mathbf{Q}_i = \mathrm{diag}(\alpha_i/h^2, \beta_i(T_1^i)^4/(16\pi^4 h^2))$, where α_i and β_i are dimensionless parameters that will be tuned to achieve the best semiactive control performance. The control gain $\mathbf{K}_i^d$ is designed using the `lqr` command in MATLAB for each pair (α_i, β_i). The commanded force is clipped to $0.15W$ to accommodate control device practical performance limits.

Model Uncertainty

While the proposed approach is capable of dealing with a wide variety of model uncertainties, this paper investigates the uncertainty in k_b and c_b, the stiffness and damping of the isolation layer, respectively. To model the disturbances, replace k_b and c_b in (8) with $k_b + \Delta k_b$ and $c_b + \Delta c_b$, respectively. By moving Δk_b and Δc_b to the right-side of the equation, (8) can be written

$$\begin{bmatrix} m_b & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{M}_s \end{bmatrix} \begin{Bmatrix} \ddot{x}_b \\ \ddot{\mathbf{x}}_s \end{Bmatrix} + \begin{bmatrix} c_b + \mathbf{r}^{\mathrm{T}}\mathbf{C}_s\mathbf{r} & -\mathbf{r}^{\mathrm{T}}\mathbf{C}_s \\ -\mathbf{C}_s\mathbf{r} & \mathbf{C}_s \end{bmatrix} \begin{Bmatrix} \dot{x}_b \\ \dot{\mathbf{x}}_s \end{Bmatrix} + \begin{bmatrix} k_b + \mathbf{r}^{\mathrm{T}}\mathbf{K}_s\mathbf{r} & -\mathbf{r}^{\mathrm{T}}\mathbf{K}_s \\ -\mathbf{K}_s\mathbf{r} & \mathbf{K}_s \end{bmatrix} \begin{Bmatrix} x_b \\ \mathbf{x}_s \end{Bmatrix}$$
$$= -\begin{bmatrix} m_b & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \mathbf{M}_s \end{bmatrix} \begin{Bmatrix} 1 \\ \mathbf{r} \end{Bmatrix} \ddot{x}_g^a + \begin{Bmatrix} 1 \\ \mathbf{0} \end{Bmatrix} u + \begin{Bmatrix} 1 \\ \mathbf{0} \end{Bmatrix} f \qquad (9)$$

where $f = -(\Delta k_b x_b + \Delta c_b \dot{x}_b)$ is a function representing the force caused by parameter uncertainties.

Generally, solving (9) results in a system of two NVIEs; *i.e.*, $\mathbf{p}(t) = \mathbf{g}(\overline{\mathbf{X}}(t))$ where $\mathbf{p}(t) = [u \; f]^{\mathrm{T}}$ and $\mathbf{g}(\overline{\mathbf{X}}(t)) = [p^d H(-p^d \dot{x}_b) \; -\Delta k_b x_b - \Delta c_b \dot{x}_b]^{\mathrm{T}}$. However, in this specific example, the influence matrices of both u and f are the same $[1 \; \mathbf{0}^{\mathrm{T}}]^{\mathrm{T}}$, so these two terms can be combined into a single NVIE equation with three parameters

$$p(t) = g(u^d, x_b, \dot{x}_b) = u^d H(-u^d \dot{x}_b) - (\Delta k_b x_b + \Delta c_b \dot{x}_b) \qquad (10)$$

Applying (10), the equation of motion (9) can be simplified to

$$\mathbf{M} \begin{Bmatrix} \ddot{x}_b \\ \ddot{\mathbf{x}}_s \end{Bmatrix} + \mathbf{C} \begin{Bmatrix} \dot{x}_b \\ \dot{\mathbf{x}}_s \end{Bmatrix} + \mathbf{K} \begin{Bmatrix} x_b \\ \mathbf{x}_s \end{Bmatrix} = -\mathbf{M} \begin{Bmatrix} 1 \\ \mathbf{r} \end{Bmatrix} \ddot{x}_g^a + \begin{Bmatrix} 1 \\ \mathbf{0} \end{Bmatrix} p \qquad (11)$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrices on the left side of (9).

Typical values for normalized standard deviation of stiffness and damping of a rubber bearing (Adachi and Unjoh, 2003; Feng et al., 2004) are adopted here: Δk_b and Δc_b are assumed to be independent, zero mean Gaussian random variables with standard deviations of $0.05k_b$ and $0.20c_b$, respectively.

Uncertainty analysis often utilizes a Monte Carlo simulation to investigate the effect of different parameter values. Herein, 10,000 random samples of stiffness and damping modifications $(\Delta k_b, \Delta c_b)$ are generated based on their corresponding distributions. A Monte Carlo simulation is used to compute the statistics of RMS and peak base drift and roof acceleration. As is shown in Table 1, the coefficients of variation of the RMS and peak responses are limited to 0.10 to 1.25 percent, which are much smaller

TABLE 1: RMS and peak of base drift and roof acceleration of a Monte Carlo simulation

Responses		Mean	Standard deviation (coeff. of variation)	
Base drift [cm]	RMS	1.34	0.00134	(0.10%)
	Peak	5.06	0.0297	(0.59%)
Roof acceleration (m/s^2)	RMS	0.66	0.00162	(0.25%)
	Peak	3.31	0.0413	(1.25%)

than the corresponding values of the parameter disturbances. This indicates that the control design is stable and the stochastic behavior of the isolation layer parameters has a very modest effect on the structural responses. Note that, the simulations are performed at the optimal design point of the semiactive controller of a deterministic system (*i.e.*, the semiactive control strategy is designed for a system with $\Delta k_b = \Delta c_b = 0$).

Computational Cost

One goal of this paper is to investigate how much reduction in computational effort can be achieved when the simulations required by Monte Carlo simulation are performed using the proposed method instead of a traditional solver.

Traditional Solver: To perform a Monte Carlo simulation for the isolation layer parameters using a traditional solver such as `ode45`, one should generate many pair of stiffness and damping disturbances $(\Delta k_b, \Delta c_b)$ based on their corresponding distributions, update the structural model for each pair, and analyze the system to determine the structural responses. This procedure is shown in Figure 1. As is demonstrated in Table 2, an `ode45` algorithm with the default relative and absolute tolerances will perform one simulation in 243.73 sec and, consequently, for a 10,000 realization Monte Carlo simulation, it will take about a month.
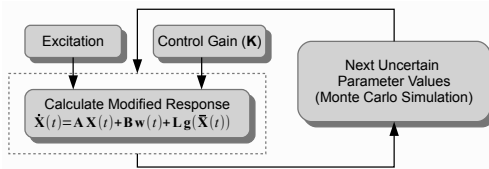
FIG. 1: Uncertainty analysis using a traditional solver.

Proposed Approach: Significant increases in computational efficiency can be achieved if the proposed approach is implemented for system model (11) because the nominal system remains unchanged for the Monte Carlo simulation; therefore, the unmodified response and the necessary impulse responses need to be calculated just once. As shown in Figure 2, each new simulation requires solving only a single NVIE equation and finding the modified response, which results in much smaller computational effort compared to the traditional solver.
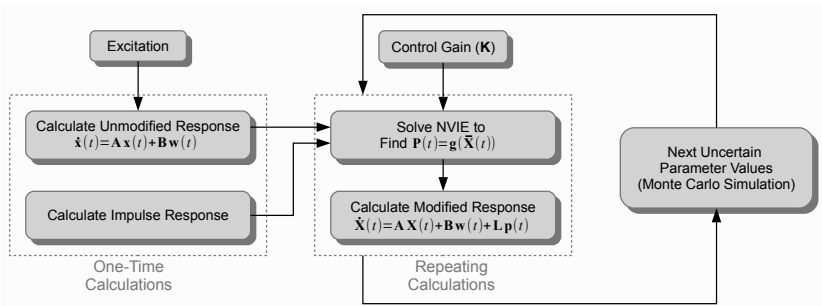
FIG. 2: Uncertainty analysis using the proposed approach.

For a fair timing comparison, both the traditional solver (*i.e.*, ode45) and the proposed method should have the same level of accuracy. Therefore, a preliminary parameter study was performed to determine the integration time step in the proposed method which will give an accuracy comparable to that of ode45 with the default tolerances (*i.e.*, relative tolerance 10^{-3} and absolute tolerance 10^{-6}); this preliminary study found that a time step Δt of about half of a millisecond gives accuracy comparable to the default ode45.

For this example, one simulation at the control design point takes 12.20 sec (8 sec for one-time calculations and 4.20 sec for the repeated calculations). However, in the case of multiple simulations, computational efficiency will be increased since each new simulation takes only another 4.20 sec for the repeated calculations. By implementing the proposed method, the 10,000 realization Monte Carlo simulation can be performed in half of a day; relative to the traditional solver, this represents a computation time ratio of about 58 (See Table 2).

TABLE 2: Cost ratio comparison.

	ode45 RelTol = 10^{-3} AbsTol = 10^{-6}	Proposed method # of steps = 2^{16} $\Delta t = 0.46\,\text{ms}$	Cost ratio
at the design point (1 simulation)	243.73 s	$8.00 + 4.20 = 12.20\,\text{s}$	20
Monte Carlo simulation (10,000 simulations)	$10^4 \times 243.73\,\text{s}$ $= 28.2$ days	$8.00 + 10^4 \times 4.20\,\text{s}$ $= 0.5$ days	58

CONCLUSIONS

The proposed method previously introduced by the authors for fast computational analysis of linear systems with local modifications has been adapted to include uncertainty analysis of a system with a smart control device. It was shown that the

equation incorporating the uncertain behavior of model parameters can be augmented by the nonlinear equation of semiactive control device. Consequently, the system of equations remains in the same form as the deterministic model and the proposed method can be implemented without any modification. The uncertainty analysis was performed using a Monte Carlo simulation for the isolation layer parameters, the rubber bearing stiffness and damping. It was shown that the control design is robust and the stochastic behavior of the isolation layer parameters has a very modest effect on the structural responses. Further, by implementing the proposed approach instead of a traditional solver such as `ode45`, the computational cost can be reduced by a factor of 58.

ACKNOWLEDGMENTS

REFERENCES

Adachi, Y. and Unjoh, S. (2003). "Seismic response characteristics of seismically isolated bridge considering hardening effect of seismic isolator." *Proceedings 10th Working Conference on Reliability and Optimization of Structural Systems*, 161–171. MAR 25-27, 2002; OSAKA, JAPAN.

Feng, D., Chen, C., Liu, W., and Tanaka, K. (2004). "A performance test study on Chinese G4 lead rubber bearings." *Proceedings 13th World Conference on Earthquake Engineering*. Paper No. 676; August 1-6, 2004; Vancouver, B.C., Canada.

Gaurav, Wojtkiewicz, S. F., and Johnson, E. A. (2011). "Efficient uncertainty quantification of dynamical systems with local nonlinearities and uncertainties." *Probabilistic Engineering Mechanics*, 26(4), 561–569.

Housner, G. W., Bergman, L. A., Caughey, T. K., Chassiakos, A. G., Claus, R. O., Masri, S. F., Skelton, R. E., Soong, T. T., Spencer, B. F., and Yao, J. T. P. (1997). "Structural control: Past, present, and future." *Journal of Engineering Mechanics*, 123(9), 897–971.

Kamalzare, M., Johnson, E. A., and Wojtkiewicz, S. F. (2013). "Computationally efficient design of optimal strategies for controllable damping devices." *Structural Control and Health Monitoring, in review*.

Skinner, R. I., Robinson, W. H., and McVerry, G. H. (1993). *An Introduction to Seismic Isolation*. Wiley, Chichester, England.

Soong, T. T. and Spencer, B. F. (2002). "Supplemental energy dissipation: state-of-the-art and state-of-the-practice." *Engineering Structures*, 24(3), 243–259.

Symans, M. D. and Constantinou, M. C. (1999). "Semi-active control systems for seismic protection of structures: a state-of-the-art review." *Engineering Structures*, 21(6), 469–487.

Optimal Semiactive Control of Elevated Highway Bridges: An Upper Bound On Performance Via a Dynamic Programming Approach

Wael M. Elhaddad and Erik A. Johnson, M. ASCE

Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 S Vermont Ave, KAP 210, Los Angeles, CA 90089; email: welhadda@usc.edu, JohnsonE@usc.edu

ABSTRACT

Controllable damping studies for seismic protection of elevated highway bridges found that a clipped LQR strategy achieves performance similar to that with a fully linear actuator when attempting to reduce the bridge deck response, but only comparable to that with a purely passive damper when seeking to reduce pier response (Erkus et al., 2002). This paper examines this more challenging pier response problem, using a dynamic programming approach to obtain the optimal dissipative control force, assuming the excitation is deterministic and known a priori, through a numerical solution of the Euler-Lagrange equations. The resulting (non-causal) optimal performance, an upper bound on the performance of any controllable damping strategy, is still rather lower than active control performance (which is causal, requiring no future excitation information). This result suggests that even highly adaptive control laws to command controllable dampers may never approach fully active system performance for some control objectives.

INTRODUCTION

Different structural control systems are used for vibration reduction in many buildings and bridges around the world (Spencer and Sain, 1997; Spencer and Nagara-jaiah, 2003). Applications of different control strategies in structures, such as passive, active, hybrid and semiactive control, were investigated extensively. Although passive control strategies (Soong and Dargush, 1997) are the best established and most well-accepted for structural vibration reduction, they are not the most efficient control technique as they cannot adapt to different loading events and structural conditions. On the other hand, active control strategies are more adaptive and efficient in vibration reduction, though there are some concerns about their reliability (Spencer and Nagara-jaiah, 2003) since a poor design can render a structure unstable and they can demand power that might not be available during an extreme loading event such as an earth-quake or severe wind exposure. Recently, semiactive strategies based on controllable passive devices have emerged as an alternative for vibration reduction in structures.

73

Controllable passive devices are ones that exert forces through purely passive means, such as a controllable damper or controllable stiffener, but have controllable properties that affect those forces. The main benefits of using semiactive control are its inherent stability, as it does not introduce energy into the controlled structure, its ability to focus on multiple (and possibly changing) objectives, exerting a force that can depend on non-local information, as well as the low power requirement that is critical in the case of natural hazards like earthquakes (Dyke et al., 1996). In addition, it has been shown that semiactive control is capable, in some cases, of achieving performance comparable to that of a fully active system (Dyke et al., 1996).

Erkus et al. (2002) investigated the effectiveness of semiactive control for seismic protection of elevated highway bridges. In their study, a semiactive control based on the clipped LQR design was employed and compared to optimally designed active and passive control techniques for different control objectives. It was found that the semiactive control can achieve performance similar to that of a fully active system when the control objective was to reduce the bridge deck response. On the other hand, when the control objective is to reduce pier response, semiactive control, designed with clipped-optimal control, can only achieve performances similar to a passive control.

Gavin and Aldemir (2005) investigated optimal semiactive control of base isolated structures subjected to ground motion. Their approach was based on a gradient based optimization algorithm. Results from a truly optimal semiactive control were compared to passive control and psuedoskyhook control. Although this optimization procedure results in the best reduction of the structural response, this approach however assumes that the ground motion is known *a priori*. For that reason, this approach represents an upper bound on the performance of causal semiactive control rules.

Alhan et al. (2006) used the same deterministic optimization approach to find optimal semiactive control forces for a single degree of freedom (SDOF) structure subjected to pulse-like ground motion. Aldemir and Gavin (2006) used a similar approach to study the effectiveness of different semiactive control strategies for structural base isolation by comparing them to the optimal semiactive control. In this paper, the same optimization procedure is used to investigate semiactive control of elevated highway bridges with an emphasis on minimizing the bridge pier response. Optimal semiactive control trajectories are evaluated; performance is compared to that of a state feedback control strategy and to its active control counterpart. The next section summarizes the methodology to be used in the subsequent numerical example.

FORMULATION

State Feedback Control Strategies using a Linear Quadratic Regulator (LQR)

Consider a structure, with n degrees of freedom, that is subjected to an external excitation and is controlled by some devices to reduce its response. The equations of motion for such a structure can be written in matrix form as:

$$\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{C}\dot{\mathbf{q}}(t) + \mathbf{K}\mathbf{q}(t) = \overline{\mathbf{B}}_{\mathbf{w}}\mathbf{w}(t) + \overline{\mathbf{B}}_{\mathbf{u}}\mathbf{u}(t) \tag{1}$$

where $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ are the mass, damping and stiffness matrices of the structural system, respectively; $\mathbf{q}$ is the displacement vector of the structure relative to the ground; $\mathbf{w}$ is the

external excitation (*e.g.*, ground acceleration or wind forces) with influence matrix $\overline{\mathbf{B}}_{\mathbf{w}}$; $\mathbf{u}$ is the damping force exerted by the control device with influence matrix $\overline{\mathbf{B}}_{\mathbf{u}}$. The same set of equations can also be written in continuous-time state space form as follows:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}_{\mathbf{w}}\mathbf{w}(t) + \mathbf{B}_{\mathbf{u}}\mathbf{u}(t) \tag{2}$$

in which

$$\mathbf{x} = \begin{Bmatrix} \mathbf{q} \\ \dot{\mathbf{q}} \end{Bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{M}^{-1}\mathbf{K} & \mathbf{M}^{-1}\mathbf{C} \end{bmatrix}, \quad \mathbf{B}_{\mathbf{w}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\overline{\mathbf{B}}_{\mathbf{w}} \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{\mathbf{u}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1}\overline{\mathbf{B}}_{\mathbf{u}} \end{bmatrix}$$

where $\mathbf{x}$ is the state vector, $\mathbf{A}$ is the state matrix, $\mathbf{B}_{\mathbf{w}}$ is the influence matrix of the excitation and $\mathbf{B}_{\mathbf{u}}$ is the influence matrix of the control force.

The optimal control design for a linear unconstrained system can be obtained in a variety of ways; one such example is using LQR, where the objective is to minimize an infinite-horizon quadratic performance index of the form:

$$J = \int_0^\infty \left[\mathbf{x}^{\mathrm{T}}\mathbf{Q}\mathbf{x} + 2\mathbf{x}^{\mathrm{T}}\mathbf{N}\mathbf{u} + \mathbf{u}^{\mathrm{T}}\mathbf{R}\mathbf{u} \right] \mathrm{d}t \tag{3}$$

in which $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{N}$ are weighting matrices for the different objective terms. If the external excitation is assumed to be white noise with zero mean, the certainty equivalence property (Van de Water and Willems, 1981) of stochastic control theory will apply, and the stochastic optimal control is equivalent to the optimal control obtained from deterministic analysis that assumes an initial condition and ignores the excitation. The LQR control design is a quadratic programming problem, whose optimal solution is a linear state feedback controller (Hespanha, 2009) with feedback gain $\mathbf{K}_{\mathrm{LQR}}$:

$$\mathbf{u}_{\mathrm{des}}(t) = -\mathbf{K}_{\mathrm{LQR}}\mathbf{x}(t) = -\left[\mathbf{R}^{-1}(\mathbf{B}^{\mathrm{T}}\mathbf{P} + \mathbf{N}^{\mathrm{T}}) \right] \mathbf{x}(t) \tag{4}$$

where $\mathbf{P}$ is the positive definite symmetric solution of the algebraic Riccati equation:

$$(\mathbf{A} - \mathbf{B}\mathbf{R}^{-1}\mathbf{N}^{\mathrm{T}})^{\mathrm{T}}\mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{B}\mathbf{R}^{-1}\mathbf{N}^{\mathrm{T}}) - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{P} + \mathbf{Q} - \mathbf{N}\mathbf{R}^{-1}\mathbf{N}^{\mathrm{T}} = 0 \tag{5}$$

Assuming the states of the structural system (*i.e.*, displacements and velocities) can be measured or estimated, then (4) can be used to determine the required control input, and the device can be commanded with the desired damping force. If active control is employed, an idealized model may be adopted with no constraints on the control force, which means the desired control force can be exerted on the structure. On the other hand, if semiactive devices are used, they can only exert dissipative forces; any commanded non-dissipative forces will not be realized. To apply desired forces with semiactive devices, the *clipped-optimal control* algorithm (Dyke

FIG. 1. Idealized model of the passivity constraint for controllable damping devices

et al., 1996) employs a primary controller that is obtained assuming unconstrained damping forces (*i.e.*, the passivity constraints of the semiactive device are ignored in this

stage), and the device will only exert the desired forces that are dissipative. This behavior can be modeled in simulations as if a secondary controller clips the nondissipative control forces.

Different semiactive devices have different constraints on the achievable damping forces; however, they all share the passivity constraint. An idealized model is adopted for the passivity constraint, as illustrated graphically in Figure 1. It is assumed that all dissipative forces are realizable by the device and all nondissipative forces are infeasible. The passivity constraint is a nonlinear constraint that can be written as:

$$u_{\text{des}} \cdot v_{\text{rel}} \geq 0 \tag{6}$$

where u_{des} is the desired control force and v_{rel} is the relative velocity across the semiactive device (chosen with sign convention such that a positive dissipative force resists motion in the positive velocity direction). The resulting control force, that the semiactive device will realize, can be calculated as follows:

$$u_{\text{sa}} = u_{\text{des}} \cdot H[u_{\text{des}} \cdot v_{\text{rel}}] = \begin{cases} u_{\text{des}}, & u_{\text{des}} \cdot v_{\text{rel}} \geq 0 \\ 0, & u_{\text{des}} \cdot v_{\text{rel}} < 0 \end{cases} \tag{7}$$

where $H[\cdot]$ is the Heaviside step function.

Optimal Control Trajectories for Deterministic Excitation

This section studies the truly optimal control assuming the excitation is known *a priori*. The approach of Gavin and Aldemir (2005) to find the optimal semiactive control of a base isolated structure subject to an earthquake ground motion is summarized.

Consider a controlled dynamical system governed by the nonlinear equation:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \tag{8}$$

for which the control objective is to minimize the integral cost function:

$$J = \int_0^{t_f} \mathscr{L}(\mathbf{x}(t), \mathbf{u}(t), t) \, dt \tag{9}$$

where $\mathscr{L}(\mathbf{x}(t), \mathbf{u}(t), t)$ is the Lagrangian of the cost function, and is a scalar function of the states and control inputs. (9) is to be minimized subject to the equality constraint represented by the system dynamics in (8). The cost function can be augmented with the equality constraints by using Lagrange multipliers so that the augmented cost function becomes:

$$J_A = J + \int_0^{t_f} \boldsymbol{\lambda}^T(t) [\mathbf{f}(\mathbf{x}, \mathbf{u}, t) - \dot{\mathbf{x}}(t)] \, dt \tag{10}$$

where $\boldsymbol{\lambda}(t)$ is a vector of Lagrange multipliers, also called the co-state vector (Kirk, 2004) (the explicit dependence of $\mathbf{x}$ and $\mathbf{u}$ on time is omitted for notational clarity). By defining the Hamiltonian as $\mathscr{H}(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t) = \mathscr{L}(\mathbf{x}, \mathbf{u}, t) + \boldsymbol{\lambda}^T(t) \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$, then (10) can be rewritten as:

$$J_A = \int_0^{t_f} \left[\mathscr{H}(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t) - \boldsymbol{\lambda}^T(t)\dot{\mathbf{x}}(t) \right] dt \tag{11}$$

Cost function (11) can be minimized by setting the first variation of J_A equal to zero (Kirk, 2004), resulting in Euler-Lagrange equations that represent the necessary optimality conditions:

$$\dot{\boldsymbol{\lambda}}(t) = \left(-\frac{\partial \mathscr{H}}{\partial \mathbf{x}} \right)^T = - \left[\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, t)}{\partial \mathbf{x}} \right]^T \boldsymbol{\lambda}(t) - \left[\frac{\partial \mathscr{L}(\mathbf{x}, \mathbf{u}, t)}{\partial \mathbf{x}} \right]^T, \quad \boldsymbol{\lambda}(t_f) = \mathbf{0} \quad (12)$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \boldsymbol{\lambda}^{\mathrm{T}}(t)\frac{\partial \mathbf{f}(\mathbf{x},\mathbf{u},t)}{\partial \mathbf{u}} + \frac{\partial \mathcal{L}(\mathbf{x},\mathbf{u},t)}{\partial \mathbf{u}} = \mathbf{0} \tag{13}$$

The Euler-Lagrange equations can be solved numerically using a steepest descent method (Kirk, 2004). Starting with an initial guess for the control trajectory, the control is updated iteratively using the Hamiltonian gradient. At each iteration, co-state dynamics in (12) are solved backward in time; the control trajectory is updated with:

$$\mathbf{u}^{k+1}(t) = \mathbf{u}^k(t) - s^k\left[\frac{\partial \mathcal{H}}{\partial \mathbf{u}}(t)\right]^{\mathrm{T}} \tag{14}$$

where $\mathbf{u}^k$ is the current control trajectory and $\mathbf{u}^{k+1}$ is the updated control trajectory after the current iteration. The scalar s^k is the step size for the current iteration. It has to be noted that the convergence rate of the optimization process is highly dependent on the proper selection of the step size at each iteration. A linear search can be carried out in the direction of the Hamiltonian gradient to determine the optimum step size for each iteration. The iterative process is repeated until the decrease in the cost function J is practically insignificant.

NUMERICAL APPLICATION TO CONTROL OF A BRIDGE MODEL

In this section, the steepest descent technique is used to find the optimal control trajectories for a bridge model, similar to the one investigated by Erkus et al. (2002). This simplified model assumes that the bridge can be represented as a two degree-of-freedom (2DOF) system, where the bridge pier and the bridge deck are each lumped into a single mass. The control device is installed between the pier and the deck as shown in Figure 2. The system parameters used for this model are $m_1 = 100\,\mathrm{Mg}$, $m_2 = 500\,\mathrm{Mg}$, $k_1 = 15.791\,\mathrm{MN/m}$, $k_2 = 7.685\,\mathrm{MN/m}$, $c_1 = 125.6\,\mathrm{kN \cdot s/m}$ and $c_2 = 0$, which results in natural frequencies $\omega_1 = 3.18\,\mathrm{rad/s}$ and $\omega_2 = 15.5\,\mathrm{rad/s}$ for the two modes, with damping ratios $\zeta_1 = 0.45\%$ and $\zeta_2 = 4.0\%$, respectively.

(a) Bridge damping system   (b) 2DOF model of the bridge
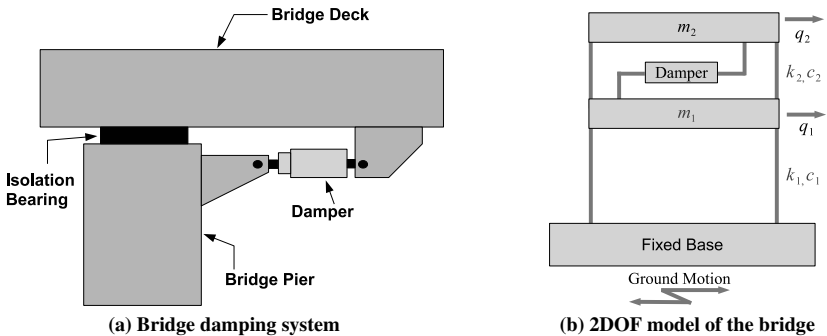
FIG. 2. Simplified bridge model

This 2DOF model can be represented in state space form (2) where:

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} c_1+c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix}, \quad \mathbf{K} = \begin{bmatrix} k_1+k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix}$$

$$\overline{\mathbf{B}}_{\mathbf{w}} = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \text{and} \quad \overline{\mathbf{B}}_{\mathbf{u}} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

The main objective of control considered in this paper is to minimize the pier response. In order to design the control for that purpose, the following quadratic cost function was selected (Erkus et al., 2002):

$$J = \int_0^\infty \left(r \left[\frac{1}{2}k_1 q_1^2 + \frac{1}{2}m_1 \dot{q}_1^2 \right] + \left[\frac{1}{2}k_2(q_2 - q_1)^2 + \frac{1}{2}m_2(\dot{q}_2 - \dot{q}_1)^2 \right] + Ru^2 \right) dt \quad (15)$$

where $r = 1000$ is selected to reflect the higher importance of reducing the pier response and R is selected to be 10^{-8} m/N, which was found in preliminary simulations to produce a controller with a high authority for vibration reduction. For LQR control design, the same objective function can be rewritten as in (3), by selecting the following coefficient matrices (Erkus et al., 2002):

$$\mathbf{Q} = \frac{1}{2} \begin{bmatrix} rk_1 + k_2 & -k_2 & 0 & 0 \\ -k_2 & k_2 & 0 & 0 \\ 0 & 0 & rm_1 + m_2 & -m_2 \\ 0 & 0 & -m_2 & m_2 \end{bmatrix} \quad \text{and } \mathbf{N} = \mathbf{0}$$

The resulting control gain in this case becomes:

$$\mathbf{K}_{\text{LQR}} = \begin{bmatrix} 939.28 & -13.37 & 65.03\,\text{s} & -35.68\,\text{s} \end{bmatrix} \text{ MN/m} \quad (16)$$

In order to find the optimal control trajectories for an active control device using the steepest descent approach, the following partial derivatives are evaluated:

$$\frac{\partial \mathbf{f}(\mathbf{x}, u, t)}{\partial \mathbf{x}} = \mathbf{A}, \quad \frac{\partial \mathbf{f}(\mathbf{x}, u, t)}{\partial u} = \mathbf{B}, \quad \frac{\partial \mathscr{L}(\mathbf{x}, u, t)}{\partial u} = 2Ru$$

$$\text{and} \quad \left[\frac{\partial \mathscr{L}(\mathbf{x}, u, t)}{\partial \mathbf{x}} \right]^{\text{T}} = \begin{bmatrix} (rk_1 + k_2)q_1 - k_2 q_2 \\ k_2 q_2 - k_2 q_1 \\ (rm_1 + m_2)\dot{q}_1 - m_2 \dot{q}_2 \\ m_2 \dot{q}_2 - m_2 \dot{q}_1 \end{bmatrix}$$

For the semiactive control case, a clipped LQR strategy is used in the simulation using the same control, by adding a secondary controller that clips the non-dissipative control force. To facilitate finding the optimal semiactive control trajectories using the steepest descent technique, the semiactive control force can be rewritten as follows:

$$u = \overline{u}H[\overline{u}](\dot{q}_2 - \dot{q}_1) \quad (17)$$

The partial derivatives involving $\overline{u}$ required for the steepest descent procedure are:

$$\frac{\partial \mathbf{f}(\mathbf{x}, \overline{u}, t)}{\partial \mathbf{x}} = \mathbf{A} + \overline{u}H[\overline{u}] \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{m_1} & \frac{1}{m_1} \\ 0 & 0 & -\frac{1}{m_2} & \frac{1}{m_2} \end{bmatrix}, \quad \frac{\partial \mathbf{f}(\mathbf{x}, \overline{u}, t)}{\partial \overline{u}} = (\dot{q}_2 - \dot{q}_1)H[\overline{u}] \begin{bmatrix} 0 \\ 0 \\ +\frac{1}{m_1} \\ -\frac{1}{m_2} \end{bmatrix},$$

$$\left[\frac{\partial \mathscr{L}(\mathbf{x}, \overline{u}, t)}{\partial \mathbf{x}} \right]^{\text{T}} = \begin{bmatrix} (rk_1 + k_2)q_1 - k_2 q_2 \\ k_2 q_2 - k_2 q_1 \\ (rm_1 + m_2)\dot{q}_1 - m_2 \dot{q}_2 - 2R\overline{u}^2(\dot{q}_2 - \dot{q}_1)H[\overline{u}] \\ m_2 \dot{q}_2 - m_2 \dot{q}_1 + 2R\overline{u}^2(\dot{q}_2 - \dot{q}_1)H[\overline{u}] \end{bmatrix}$$

$$\text{and} \quad \frac{\partial \mathscr{L}(\mathbf{x}, \overline{u}, t)}{\partial \overline{u}} = 2R\overline{u}(\dot{q}_2 - \dot{q}_1)^2 H[\overline{u}]$$

Simulations were carried out for the bridge model using state feedback control for both active and semiactive control. In addition, the steepest descent optimization approach was used to determine the optimal control trajectories for both cases. The earthquake considered in this analysis is the 1940 El Centro record. A total of 2000 iterations were carried out to ensure convergence of the optimization process, however Figure 3 shows the cost function for the first

FIG. 3. Convergence of the cost functions during iterations

50 iterations only. Table 1 shows the comparison of results for both active and semiactive control. In this table, the part of the cost that represents the integral of the term $\mathbf{x}^T\mathbf{Q}\mathbf{x}$ is denoted by $J_\mathbf{x}$ while that for the term $\mathbf{u}^T\mathbf{R}\mathbf{u}$ is denoted by $J_\mathbf{u}$.

TABLE 1. Comparison of peak pier responses and cost values for state feedback control and optimal control using El Centro earthquake

	Active control			Semiactive control		
	LQR	Optimal	Δ [%]	CLQR	Optimal	Δ [%]
$q_1^{\max}$ [cm]	1.25	0.39	−69	4.50	2.56	−43
$J_\mathbf{x}$ [MJ · s]	3.72	0.587	−84	25.17	12.37	−51
$J_\mathbf{u}$ [kJ · s]	29.05	9.058	−69	12.41	4.654	−63
J [MJ · s]	3.75	0.596	−84	25.18	12.38	−51

Note: Δ is the percent change from the LQR-based state feedback control to the truly optimal control; negative numbers are improvements.

Figure 4 shows the pier response during the first 10 seconds of El Centro earthquake for both active and semiactive control. In both cases, the optimal control trajectories can reduce the response significantly compared to the state feedback control laws. It is also evident that active control performance is much better than semiactive control for the considered objective function; this was shown previously by Erkus et al. (2002).

(a) Active control   **(b) Semiactive control**
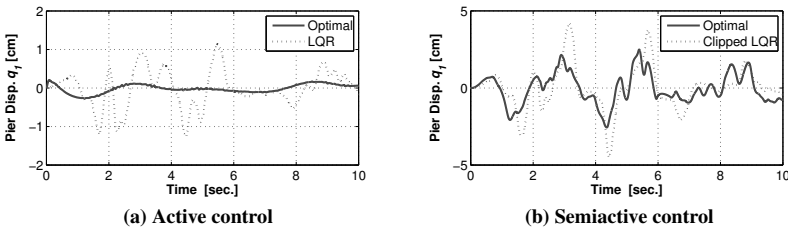
FIG. 4. Pier response during the first 10 seconds of El Centro earthquake

Although the values in Table 1 show that the optimal semiactive control can reduce the pier displacement significantly to 2.56 cm (43% reduction compared to clipped LQR) , it also shows that this displacement is much higher than that achieved by

the active system (1.254 cm for LQR). This in fact means that even a highly adaptive semiactive control laws may not be able to rise to the performance of the active control for the considered control objective. It has to be noted here that the performance of the optimal semiactive control trajectories is considered an upper bound for any adaptive semiactive control law that can incorporate ground excitation information.

CONCLUSIONS

Based on the results discussed in the previous section for the bridge model, the following conclusions can be drawn:

1. Adaptive control strategies for elevated highway bridges have significant room for improving the control performance, assuming additional information about the ground excitation is available and can be incorporated in the design of the control strategy.
2. The improvement of adaptive semiactive control strategies might be limited for some control objectives, if compared to the improvement that adaptive active control strategies can achieve.
3. For the minimization of bridge pier response, the performance of optimal semiactive control cannot rise to the performance of optimal active control, or even the state feedback active control based on LQR design.

REFERENCES

Aldemir, U. and Gavin, H. (2006). "Optimal semiactive control of structures with isolated base." *International Applied Mechanics*, 42, 235–240.

Alhan, C., Gavin, H., and Aldemir, U. (2006). "Optimal control: Basis for performance comparison of passive and semiactive isolation systems." *Journal of Engineering Mechanics*, 132(7), 705–713.

Dyke, S. J., Spencer, B. F., Sain, M. K., and Carlson, J. D. (1996). "Modeling and control of magnetorheological dampers for seismic response reduction." *Smart Materials and Structures*, 5(5), 565–575.

Erkus, B., Abé, M., and Fujino, Y. (2002). "Investigation of semi-active control for seismic protection of elevated highway bridges." *Engineering Structures*, 24(3), 281–293.

Gavin, H. and Aldemir, U. (2005). "Optimal control of earthquake response using semiactive isolation." *Journal of Engineering Mechanics*, 131(8), 769–776.

Hespanha, J. P. (2009). *Linear Systems Theory*. Princeton Press, Princeton, New Jersey.

Kirk, D. E. (2004). *Optimal Control Theory: An Introduction*. Dover Publications.

Soong, T. T. and Dargush, G. F. (1997). *Passive Energy Dissipation Systems in Structural Engineering*. Wiley, Chichester.

Spencer, Jr., B. F. and Nagarajaiah, S. (2003). "State of the art of structural control." *Journal of Structural Engineering*, 129(7), 845–856.

Spencer, Jr., B. F. and Sain, M. K. (1997). "Controlling buildings: a new frontier in feedback." *Control Systems, IEEE*, 17(6), 19–35.

Van de Water, H. and Willems, J. (1981). "The certainty equivalence property in stochastic control theory." *IEEE Transactions on Automatic Control*, 26(5), 1080–1087.

Control for Multiple Objectives in Different Magnitude Excitations

E. Hemmat-Abiri and E.A. Johnson, M. ASCE

Sonny Astani Department of Civil and Environmental Engineering
University of Southern California, 3620 S Vermont Ave, Los Angeles, CA 90089
email: hemmatab@usc.edu, JohnsonE@usc.edu

ABSTRACT

Linear structural control strategies generally cannot focus on different objectives in different excitations. Yet, this is a goal that has real meaning in many structural control applications. For example, minimizing structural drift is necessary during strong earthquakes to mitigate damage to the structure; however, occupant comfort and the safety of building contents and nonstructural components in the much-more-frequent moderate earthquakes demand reductions in absolute accelerations. Drift and acceleration reduction in different excitations are competing objectives that cannot be achieved with conventional linear strategies. Herein, a control strategy is presented that achieves significant improvements in both serviceability (acceleration mitigation) in weak/moderate excitations and life safety (drift mitigation) in extreme events, all in one general control law. A numerical study of a simple single degree-of-freedom system, using scaled records of several historical earthquakes, controlled with conventional LQR approaches and the proposed control, demonstrates that the proposed control law successfully accomplishes both goals.

INTRODUCTION

Each year, excessive loading produced by extreme events, such as earthquakes and hurricanes, causes damage and failure of many buildings and infrastructures, with losses to public and private building owners in terms of actual structural damage and its economic impacts as well as injury and loss of life. Moreover,
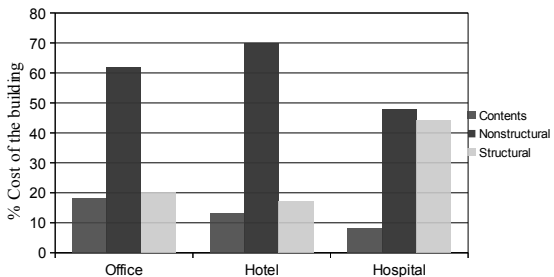
FIG. 1: Cost comparison of building components (adapted from Taghavi and Miranda (2003))

moderate-to-strong winds and weak-to-moderate earthquakes occur much more fre-
quently, affecting human comfort and causing damage to nonstructural components and
building contents in addition to loss of occupancy. Damage to building nonstructural
components and other building contents plays a very important role in the required repair
cost and time; together, these represent the vast majority of the cost of commercial struc-
tures, from 65–85% in many cases (Taghavi and Miranda, 2003) as shown in Figure 1.
Nonstructural components are vulnerable to large accelerations, with damage initiating
at low levels of external excitations; even though a weak earthquake may cause little
(if any) structural damage due to the small induced structural drifts, the accelerations
can easily result in significant damage to nonstructural components such as ceilings,
piping and so forth (Taghavi and Miranda, 2003). Similarly, common wind-induced
vibration may not approach a level that may affect structural safety but may still cause
significant discomfort to the occupants and nonstructural effects; a common reminder
is the 60-story John Hancock Tower in Boston, which experienced acceleration levels
far beyond those intended for occupant comfort, and resulting in falling panels of glass
from the many windows on the tower (Christenson, 2002; Williams and Kareem, 2003;
Schwartz, 2004).

The most critical building response in the case of severe excitation — such
as strong earthquakes, hurricanes and other extreme dynamic loads — is structural
drift; exceeding drift design values may result in severe damage to the structure and
its occupants and contents. In more moderate excitation, which occurs much more
frequently, the absolute accelerations are more likely to directly affect serviceability,
both in terms of occupant comfort and the safety and continued operation of building
contents. Many structural control systems implemented to date are designed more for
safety because it is often difficult, to simultaneously design for extreme events.

This paper gives an example of a control strategy that can achieve both service-
level and life safety-level objectives with one general control law. This particular control
strategy is not intended to be optimal, but to show that a strategy exists to reduce
drift in the rare extreme events and reduce absolute acceleration in the more frequent
weak-to-moderate ground motions.

Figure 2 shows a typical
active control block diagram of
the flow of signals and informa-
tion within a structural control
system. The control strategy will
determine the control force based
on the feedback data of structural
response and feedforward of ex-
citation data.
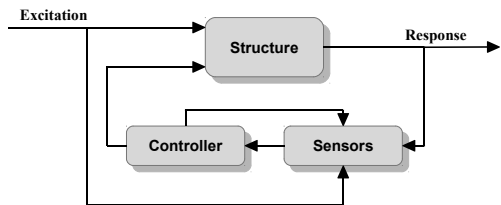
FIG. 2: Control block diagram

CONTROLLER TO ACHIEVE SERVICEABILITY AND LIFE SAFETY GOALS

Consider an n degree-of-freedom (DOF) linear system subjected to external
disturbance **w**. The equations of motion, and those governing sensor measurements and

responses, can be expressed in state-space form

$$\dot{\mathbf{q}} = \mathbf{A}\mathbf{q} + \mathbf{B}\mathbf{u} + \mathbf{E}\mathbf{w}$$
$$\mathbf{y} = \mathbf{C_y}\mathbf{q} + \mathbf{D_y}\mathbf{u} + \mathbf{F_y}\mathbf{w} + \mathbf{v} \qquad (1)$$
$$\mathbf{z} = \mathbf{C_z}\mathbf{q} + \mathbf{D_z}\mathbf{u} + \mathbf{F_z}\mathbf{w}$$

where state vector $\mathbf{q}$ contains generalized displacements and velocities; sensor measurement vector $\mathbf{y}$ may contain absolute accelerations, relative drifts and/or strains; $\mathbf{z}$ is a vector of outputs to be regulated (typically interstory drifts, absolute accelerations and internal forces); $\mathbf{u}$ is a vector of control forces; $\mathbf{v}$ is the sensor noise vector; state matrix $\mathbf{A}$, input matrices $\mathbf{B}$ and $\mathbf{E}$, output matrices $\mathbf{C_y}$ and $\mathbf{C_z}$, and feedthrough matrices $\mathbf{D_y}$, $\mathbf{D_z}$, $\mathbf{F_y}$ and $\mathbf{F_z}$ all depend on the usual structural system matrices, the force locations and orientations, etc.

To achieve the two different objectives, drift reduction and absolute acceleration reduction, in different levels of excitations, define the following two cost functions that should be minimized:

$$J_{\text{drift}} = \int_0^{t_f} [\mathbf{z}^T\mathbf{Q_d}\mathbf{z} + \mathbf{u_d}^T\mathbf{R}\mathbf{u_d}]dt \qquad \text{and} \qquad J_{\text{accel}} = \int_0^{t_f} [\mathbf{z}^T\mathbf{Q_a}\mathbf{z} + \mathbf{u_a}^T\mathbf{R}\mathbf{u_a}]dt$$

in which output vector $\mathbf{z} = [\Delta\mathbf{x}^T \quad (\ddot{\mathbf{x}}^a)^T]^T$ contains drift $\Delta\mathbf{x}$ and absolute accelerations $\ddot{\mathbf{x}}^a$; weighting matrices $\mathbf{Q_d} = [[\mathbf{q_d}\ \mathbf{0}]^T \quad [\mathbf{0}\ \mathbf{0}]^T]$ and $\mathbf{Q_a} = [[\mathbf{0}\ \mathbf{0}]^T \quad [\mathbf{0}\ \mathbf{q_a}]^T]$ force the cost functions to focus on drift and acceleration, respectively; and $\mathbf{q_d}$, $\mathbf{q_a}$ and $\mathbf{R}$ are weighting coefficients on structural drift, absolute accelerations and control forces.

Linear control minimizing one or the other of the two cost functions J_{drift} and J_{accel}, which are quadratic in states and the control force (the latter is necessary for the optimization to be well-posed), can be accomplished for an infinite time horizon by using MATLAB's `lqr` command. Each of these objective functions will result in a different optimal linear control law, namely $\mathbf{u_d} = -\mathbf{K_d}\mathbf{q}$ and $\mathbf{u_a} = -\mathbf{K_a}\mathbf{q}$ in the optimal linear control gains $\mathbf{K_d}$ and $\mathbf{K_a}$. However, as $\mathbf{u_d}$ is designed to minimize the drift, it may result in increased accelerations, which would be undesirable in the moderate excitations. Similarly, $\mathbf{u_a}$ will achieve reductions in acceleration, but is not optimal in reducing drift in the case of strong excitation.

There are a multitude of possible control laws that may satisfy the two objectives in their corresponding excitation levels. The point of this paper is to show that there does indeed exist a control strategy that satisfies the goals. This specific control strategy uses a simple combination of the two optimal linear control forces $\mathbf{u_a}$ and $\mathbf{u_d}$:

$$\mathbf{u} = \gamma(t,\mathbf{q},\mathbf{w})\mathbf{u_d} + [1 - \gamma(t,\mathbf{q},\mathbf{w})]\mathbf{u_a} = -\{\mathbf{K_d}\gamma(t,\mathbf{q},\mathbf{w}) - \mathbf{K_a}[1 - \gamma(t,\mathbf{q},\mathbf{w})]\}\mathbf{q} \qquad (2)$$

The weighting coefficient γ, which may depend on time, structural response and excitation, should be determined appropriately such that the time-varying control in (2) approaches $\mathbf{u_d}$ when the excitation is strong and closely tracks $\mathbf{u_a}$ under weak/moderate excitation. There are many possible forms for γ that can satisfy these objectives. One is $\gamma = \Gamma^4/[\Gamma^4 + \alpha]$, where α is a parameter that adjusts the transition from $\mathbf{u_a}$ to $\mathbf{u_d}$, and

$$\Gamma(t) = \max_{\tau \leq t} \|\mathbf{w}(\tau)\| \qquad (3)$$

which is the running peak of some norm of the excitation; for example, in an earthquake, $\Gamma(t)$ is the peak ground acceleration seen up through time t if $\|\mathbf{w}(\tau)\| = |\ddot{x}_g(\tau)|$ for ground acceleration $\ddot{x}_g(t)$. It should be noted that Γ increases, or stays the same, over the duration of the excitation. Then, (2) can be rewritten

$$\mathbf{u} = -[\mathbf{K}_d\Gamma^4(t) + \mathbf{K}_a\alpha]\mathbf{q}/[\Gamma^4(t) + \alpha] \qquad (4)$$

NUMERICAL EXAMPLE

To demonstrate the potential of a control strategy to achieve both goals, a numerical simulation is performed using control law (4) for a single degree of freedom (SDOF) building structure with parameters $m = 29485\,\text{kg}$, $c = 23710\,\text{N·s/m}$ and $k = 1.1640\,\text{MN/m}$, which are adapted from Ramallo et al. (2002) (who based their SDOF model on a 5DOF model from Kelly et al. (1987)), but with the stiffness reduced so that the structure has a 1 Hz natural frequency (the resulting damping is 6.4%). The weighting coefficients in J_{drift} and J_{accel} are selected such that the cost functions are dimensionless: $q_d = 10\,\text{m}^{-2}$, $q_a = 10(m/k)^2 = 0.0064\,\text{s}^4/\text{m}^2$ and $R = 1/k^2 = 0.73804\,(\text{MN})^{-2}$. The strong and weak excitations are chosen to be 3.0- and 0.1-scaled versions of the N-S component of the 1940 El Centro earthquake, with peak ground accelerations (PGAs) of $10.251\,\text{m/s}^2$ and $0.3417\,\text{m/s}^2$, respectively. A preliminary parameter study identified $\alpha = 50\,\text{m}^2/\text{s}^4$ met the objectives. The control gains $\mathbf{K}_a$ and $\mathbf{K}_d$ for the linear strategies are computed using MATLAB's lqr command. The responses were computed, using MATLAB's Simulink, for the time-varying control law and for linear control laws denoted LQR$_{\text{drift}}$ and LQR$_{\text{accel}}$. The responses to the weak excitation are shown in Figure 3, and to the strong excitation in Figure 4; peak and root-mean-square (RMS) metrics of drift, acceleration and control force are listed for both excitations in Table 1.

Clearly, the time-varying control strategy closely tracks the LQR$_{\text{drift}}$ design in the case of strong excitation and it performs almost the same as LQR$_{\text{accel}}$ in the weak excitation, demonstrating that the same time-varying control law can achieve both of the competing objectives. Table 1 shows that, in the weak excitation (0.1×El Centro), the maximum absolute acceleration response controlled with the time-varying strategy is equal to $0.1236\,\text{m/s}^2$, which is nearly identical to the corresponding $0.1237\,\text{m/s}^2$ value with the LQR$_{\text{accel}}$ design; but the drift-focused LQR$_{\text{drift}}$ has much larger acceleration of $0.3896\,\text{m/s}^2$. On the other hand, for the large excitation (3.0×El Centro), the maximum

TABLE 1: Performance of time-varying control under scaled El Centro excitations

	0.1 El Centro PGA = 0.3417 m/s^2			3 El Centro PGA = 10.251 m/s^2		
	LQR$_{\text{drift}}$	LQR$_{\text{accel}}$	Time-Varying	LQR$_{\text{drift}}$	LQR$_{\text{accel}}$	Time-Varying
Peak displ. [cm]	0.1938	0.5501	0.5497	5.8131	16.5019	5.8444
RMS displ. [cm]	0.0250	0.1014	0.1013	0.7489	3.0432	0.8083
Peak abs. accel. [m/s^2]	0.3896	0.1236	0.1237	11.6885	3.7077	11.7483
RMS abs. accel. [m/s^2]	0.0468	0.0228	0.0228	1.4047	0.6850	1.3982
Peak control force / mg	0.0327	0.0179	0.0179	0.9798	0.5382	0.9855
RMS control force / mg	0.0040	0.0033	0.0033	0.1191	0.0979	0.1182
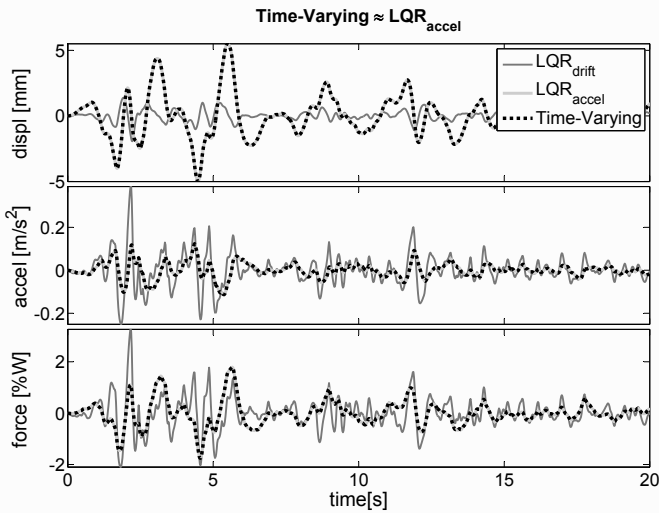
FIG. 3: Controlled response comparison for 0.1×El Centro
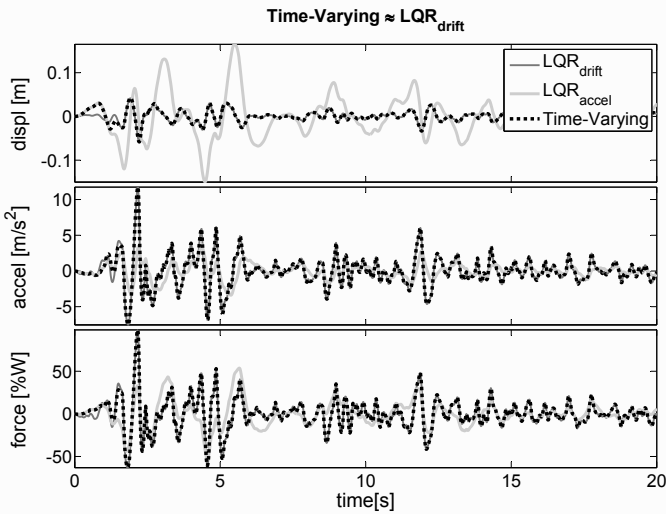
FIG. 4: Controlled response comparison for 3.0×El Centro

TABLE 2: Performance of time-varying control under scaled 1994 Northridge excitations

	0.1 Northridge PGA = 0.3613 m/s²			3 Northridge PGA = 10.8405 m/s²		
	LQR_{drift}	LQR_{accel}	Time-Varying	LQR_{drift}	LQR_{accel}	Time-Varying
Peak displ. [cm]	0.1350	0.5382	0.5380	4.0486	16.1463	4.0919
RMS displ. [cm]	0.0326	0.1668	0.1667	0.9781	5.0048	0.9882
Peak abs. accel. [m/s²]	0.2592	0.1425	0.1427	7.7771	4.2755	7.7285
RMS abs. accel. [m/s²]	0.0614	0.0356	0.0357	1.8413	1.0688	1.8324
Peak control force / mg	0.0223	0.0176	0.0176	0.6680	0.5279	0.6617
RMS control force / mg	0.0052	0.0053	0.0053	0.1562	0.1583	0.1552

drifts with both the time-varying and LQR_{drift} strategies are about 5.8 cm, whereas the drift with the LQR_{accel} control is much larger at 16.50 cm. The same trend exists in the maximum and RMS values of control force and other metrics.

The control law is optimized for scaled El Centro excitations, however it should perform well under other excitations as well. Figures 5 and 6 and Table 2 shows the performance of the time-varying control along with the two optimal linear control laws for 0.1 and 3.0-scaled 1994 Northridge excitations. Also, to investigate the performance of time-varying control under moderate to strong excitation, the system is subjected to the 1994 Northridge and 1995 Kobe excitation, with results shown in Table 3.

Clearly, the time-varying control focuses on acceleration reduction for weak excitation 0.1-scaled Northridge. For moderate excitation Northridge, it reduces both drift and acceleration at some acceptable low level, and it reduces drift under strong excitation 3.0-scaled Northridge and Kobe. In fact, the drift reduction provided by the time-varying control is much more significant compared with that of LQR_{drift} for 3.0-scaled Northridge excitation. It must be noted that force saturation was not introduced into this example for the sake of simplicity; any real implementation would, of course, need to use lower force levels, perhaps on the order of 10–15% of structure weight, possibly by introducing a saturation in the commanded force to accommodate control device capacity, distribution of the load into the structure and so forth.

TABLE 3: Performance of time-varying control under 1994 Northridge and 1995 Kobe excitations

	Northridge PGA = 3.613 m/s²			Kobe PGA = 8.1782 m/s²		
	LQR_{drift}	LQR_{accel}	Time-Varying	LQR_{drift}	LQR_{accel}	Time-Varying
Peak displ. [cm]	1.3495	5.3821	2.2387	5.2656	11.5746	5.3370
RMS displ. [cm]	0.3260	1.6683	0.4254	0.6307	2.0892	0.6605
Peak abs. accel. [m/s²]	2.5924	1.4252	2.2368	9.6192	3.6458	9.6747
RMS abs. accel. [m/s²]	0.6138	0.3563	0.5626	1.1765	0.5439	1.1849
Peak control force / mg	0.2227	0.1760	0.1730	0.7891	0.4567	0.7925
RMS control force / mg	0.0521	0.0528	0.0447	0.0996	0.0708	0.0998
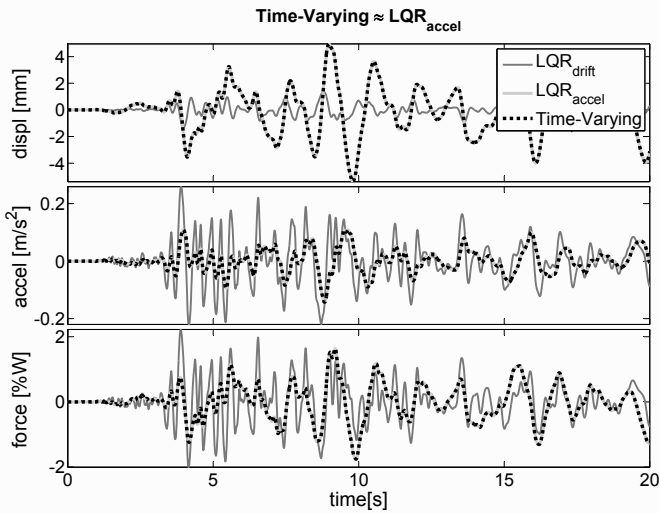
FIG. 5: Controlled response comparison for $0.1 \times$ Northridge
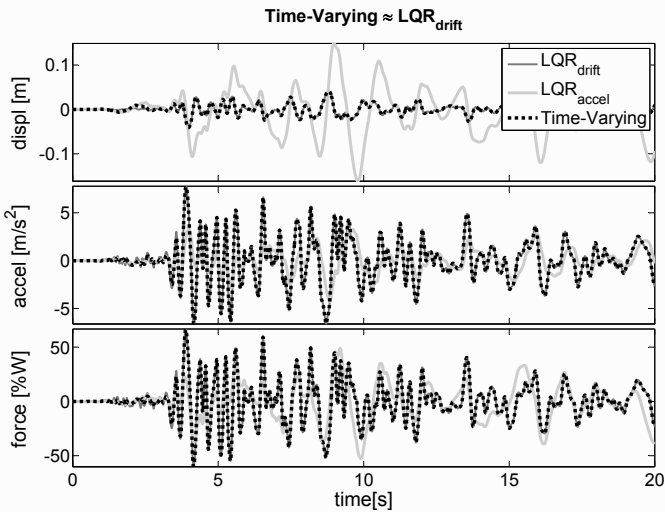
FIG. 6: Controlled response comparison for $3.0 \times$ Northridge

CONCLUSIONS

It is imperative for the civil engineering community to develop and implement strategies to mitigate structural responses in both weak and strong excitations, reducing the absolute acceleration responses in weak-to-moderate excitations to prevent nonstructural and building content damage and ensure occupant comfort, and decreasing the drift responses in strong excitations to avoid structural damage and possible structural collapse — for both newly designed structures and retrofitting of existing structures.

This paper presents an example of a control strategy that can achieve both service-level and life safety-level objectives with one general control law. This particular control strategy is not intended to be optimal, but to show that a strategy exists to reduce drift in the rare extreme events and reduce absolute acceleration in the more frequent weak to moderate ground motions.

The performance of the time-varying control strategy is demonstrated by intensive numerical simulations for a SDOF building model subjected to different excitations, showing that it can successfully achieve both life safety (drift reduction) and serviceability (acceleration reduction) objectives at different excitation levels. However, it is a simple unoptimized example. Further study and investigation is needed for deriving and optimizing time-varying control strategies that achieve both objectives.

REFERENCES

Christenson, R. E. (2002). "Semiactive control of civil structures for natural hazard mitigation: Analytical and experimental studies." Ph.D. thesis, Univ. of Notre Dame, Indiana, United States.

Kelly, J. M., Leitmann, G., and Doyle, A. G. (1987). "Robust control of base-isolated structures under earthquake excitation." *Journal of Optimization Theory and Applications*, 53, 159–180.

Ramallo, J. C., Johnson, E. A., and Spencer, Jr., B. F. (2002). "'Smart' base isolation systems." *Journal of Engineering Mechanics*, 128(10), 1088–1099.

Schwartz, T. A. (2004). "The John Hancock tower glass failure - debunking the myths." *Structures Congress 2000: Advanced Technology in Structural Engineering*, Philadelphia, PA, United States.

Taghavi, S. and Miranda, E. A. (2003). "Response assessment of nonstructural building elements." *Report No. 5*, Pacific Earthquake Engineering Research Center, PEER.

Williams, T. and Kareem, A. (2003). "Performance of building cladding in urban environments under extreme winds." *Proceedings of 11th International Conference on Wind Engineering*, Lubbock, TX.

Improving Substructure Identification using Structural Control with Limited Sensor Measurements

C. DeVore, S.M. ASCE, and E.A. Johnson, M. ASCE

Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 S Vermont Ave, Los Angeles, CA 90089; email: cdevore@usc.edu, JohnsonE@usc.edu

ABSTRACT

Substructure identification is a structural health monitoring technique that identifies a reduced order model of structural behavior using local vibration measurements. Previous research by the authors on this inherently decentralized technique has shown that substructure identification can detect and localize stiffness changes in a shear building. However, identification performance varies from story to story, with a few not satisfactorily identified. To overcome this limitation, Zhang and Johnson (2012c) showed how to use a structural control device to temporarily change the dynamics of the structure to improve identification performance. However, this method assumes full state feedback which is not achievable in full-scale implementations. This paper investigates structural control using limited sensor measurements to improve substructure identification. A 12-ft four-story single-bay steel structure model, subject to low levels of ground motion, is used as the test-bed for numerical simulations. Different combinations of acceleration sensor locations are compared to understand the performance tradeoffs.

INTRODUCTION

The past two decades have seen an increase in the number of researchers using substructure identification to perform structural health monitoring (SHM) of civil structures. The first use was Koh et al. (1991), using an extended Kalman filter (EKF) to identify the story stiffness and damping parameters in shear building. Other studies include Xing and Mita (2012) who proposed a time domain substructure identification procedure that utilizes an auto-regressive moving average with exogenous input (ARMAX) model. Hou et al. (2011) developed a substructure isolation method using the virtual distortion method. Yuen and Katafygiotis (2006) presented a substructure identification procedure in a Bayesian context.

Zhang and Johnson (2012a,b) proposed a substructure identification method using a specifically constructed frequency-domain substructure estimator. This choice allows the analyst to use long-stationary time histories to compute averaged frequency

response quantities that admit desirable statistical properties. The identification performance of the developed estimator was found to vary between substructures and sometimes resulted in mis-identification. To overcome this result, Zhang and Johnson (2012c) proposed temporarily re-purposing a structural control device to make deterministic changes to the structure's dynamics in such a way that improves identification performance. This procedure was found to be very effective at improving identification performance in the presence of measurement noise.

A previous study (DeVore et al., 2012) used substructure identification to identify story-level stiffness and damping in a 12 ft., four-story, steel structure subjected to low levels of (band-limited) white noise base motion; stationary responses were measured by accelerometers installed at each floor. Several damage scenarios were tested where the story stiffness of one floor was decreased by loosening connections. Damage was successfully detected in two stories; however, the second story was not successfully identified.

This study will use a structural control device to overcome the poor substructure identification performance of the second story found in the results of DeVore et al. (2012). Numerical simulation using the nominal parameters will be used to create an equivalent reduced order model (ROM). Then a full-state feedback controller will be designed. The controller will be implemented using an optimal observer designed for various sensor configurations. Results are provided showing the predicted performance for each arrangement and the damage detection performance found using Monte Carlo simulation (MCS).

SUBSTRUCTURE IDENTIFICATION

This section will demonstrate the steps necessary to formulate an estimator suitable for substructure identification. While the presentation is different, the derived chain structure estimator is based on Zhang and Johnson (2012b).

Estimator Formulation

To identify the stiffness and damping of the ith story (k_i and c_i), start with the equation of motion (EOM),

$$m_i \ddot{x}_i + c_i(\dot{x}_i - \dot{x}_{i-1}) + k_i(x_i - x_{i-1}) + c_{i+1}(\dot{x}_i - \dot{x}_{i+1}) + k_{i+1}(x_i - x_{i+1}) = 0 \quad (1)$$

where x_i is the displacement of the floor level with respect to an inertial reference frame.

The next step is to form the model function H_{MOD}, which is a reduced order representation of the local behavior of the substructure. A natural choice for H_{MOD} is a single degree of freedom (SDOF) oscillator because its dynamics are simple and can be characterized with only a few parameters. For the shear building model, the model function is,

$$H_{\text{MOD}}(s) = \frac{1}{1 + \frac{c_i}{m_i s} + \frac{k_i}{m_i s^2}} \quad (2)$$

Once the model function is selected, it is necessary to manipulate the EOM to generate an equation that isolates the model function. First, take the Laplace transform of EOM (1); then, algebraically manipulate the transformed equation to isolate the chosen model function on one side of the equation. Initial conditions can be neglected because signal

processing will use a tapered window function and the displacement and velocity can be rewritten as the acceleration. The final step is to rewrite the Laplace transform of the response variables in terms of their transfer functions $H_{\ddot{x}_i, \ddot{u}_g}(s) = \ddot{X}_i(s)/\ddot{U}_g(s)$. In this example, ground motion $\ddot{u}_g$ is taken as the excitation (though the excitation could take other forms). After sufficient algebraic manipulation, (1) becomes:

$$\frac{1}{1 + \frac{c_i}{m_i s} + \frac{k_i}{m_i s^2}} = \frac{H_{\ddot{x}_i, \ddot{u}_g}(s) - H_{\ddot{x}_{i-1}, \ddot{u}_g}(s)}{H_{\ddot{x}_{i-1}, \ddot{u}_g}(s) + [H_{\ddot{x}_{i+1}, \ddot{u}_g}(s) - H_{\ddot{x}_i, \ddot{u}_g}(s)]\left(\frac{c_{i+1}}{m_i s} + \frac{k_{i+1}}{m_i s^2}\right)} \tag{3}$$

By examining (3), a substructure identification estimator has been successfully constructed. On the left side is the model function $H_{\text{MOD}}(s)$; on the right side is a function, denoted the *function of estimated quantities $H_{\text{EST}}(s)$*, that maps the measured responses and *a priori* parameters.

To identify the optimal estimates of the stiffness and damping parameters, the identification functional is formed to find the distance between $H_{\text{MOD}}(s)$ and $H_{\text{EST}}(s)$.

$$\min_{[k_i, c_i]} \int |H_{\text{MOD}}(s) - H_{\text{EST}}(s)|^2 ds \tag{4}$$

Using (4), different values of k_i and c_i are tested to find the optimal value corresponding to the minimum distance between $H_{\text{MOD}}(s)$ and $H_{\text{EST}}(s)$.

The major benefit of substructure identification is that measured responses, with complicated dynamics, are simplified through the mapping of the function of estimated quantities. The individual transfer functions of the response variables have complicated dynamics with multiple poles and zeros. However, when they are combined through the function of estimated quantities, their combined behavior simplifies into that of a SDOF oscillator.

Error Prediction

Substructure identification is useful for detecting component-level damage. However, there are certain structure–substructure combinations that create difficulty for analysis. These difficulties occur when there is not sufficient response consistent with the substructure model. In DeVore et al. (2012), significant identification error is predicted when there are low levels of interstory acceleration response within the frequency bandwidth of the substructure model function $H_{\text{MOD}}(s)$. This study uses the weighted integral interstory acceleration response as a predictor of substructure identification performance. Other similar prediction functions were considered in Zhang and Johnson (2012c).

CONTROLLED SUBSTRUCTURE IDENTIFICATION

A full state feedback controller is designed. First, the operator is defined for a full state controller allowing for a consistent state space description. Next, the control design problem is defined in terms of the poles of the closed loop system. Following, linear constraints on the search space are used to meet stability and damping targets while nonlinear constraints are used to ensure realistic performance.
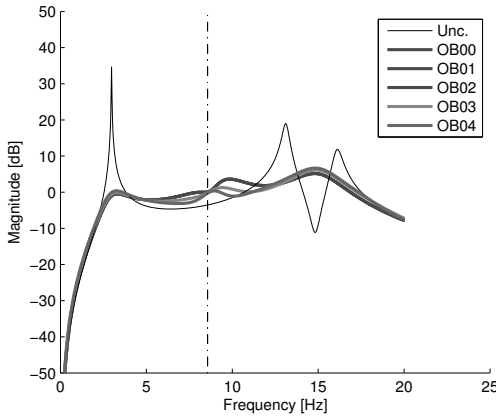
FIG. 1: Second Story Interstory Acceleration Response for various observers with substructure natural frequency shown as a dashed line ($\omega_0 = \sqrt{k_i/m_i}$). Nominal parameters are used.

Full State Feedback Controller

A full state feedback controller is defined by multiplying a feedback gain matrix **K** by the state vector and applying the resulting signal as a control force $\mathbf{u}_c = -\mathbf{K}\mathbf{x}$. By convention, feedback control is applied as negative feedback and that convention is respected within this work.

Using the feedback gain matrix, the controlled system can be defined as a linear operator,

$$T^c := \begin{cases} \dot{\mathbf{x}} = (\mathbf{A} - \mathbf{BK})\mathbf{x} + \mathbf{G}\ddot{u}_g \\ \mathbf{y} = (\mathbf{C} - \mathbf{DK})\mathbf{x} + \mathbf{H}\ddot{u}_g + \mathbf{v} \end{cases} \tag{5}$$

The state space matrices are defined for a shear building with an ideal active mass driver (AMD) installed on the roof.

This study will design a controller for an AMD installed on the top floor. It is assumed that the secondary controller of the AMD has high-authority so that the control device has static unity gain. This assumption is unlikely to be true in practice, but further considerations are outside the scope of this study.

Control Design

The control gain matrix is designed using nonlinear optimization subject to linear and nonlinear constraints. A nonlinear objective functional is calculated for each sample point and the optimization algorithm continues until a local minimum is encountered.

The objective functional is selected to increase the interstory acceleration response of the closed loop system within the identification bandwidth. Therefore, the functional J takes the form,

$$J(\mathbf{q}) := \int_{\omega_l}^{\omega_u} \left| W(j\omega) \left(T_i^c(j\omega; \mathbf{q}) - T_{i-1}^c(j\omega; \mathbf{q}) \right) \right| d\omega \tag{6}$$

where $W(j\omega)$ is the weighting function taken to be the model function H_{MOD}; ω_l and ω_u are the lower and upper bounds of the identification frequency bandwidth, respectively; and $T_i^c(j\omega; \mathbf{q})$ is the transfer function of the closed loop operator for a specific control gain evaluated at the frequency ω for the ith floor acceleration output. J is understood to be the weighted integral of the interstory acceleration response.

With the state space matrices known, J can vary with different choices of $\mathbf{K}$. Previous studies (Zhang and Johnson, 2012c) used the various entries of the control gain matrix as the search space. Herein, the search space $\mathbf{q} \in \mathbb{R}^{2N}$ is the set of poles of the closed loop system. The search space has a one-to-one correspondence with the various control gain entries while providing unique computational benefits by specifying some constraints as linear constraints.

The initial search vector, $\mathbf{q}_0$, is selected to be the uncontrolled poles of the structure. This is found to have a stabilizing effect on control design and yields a consistent controller. The final solution $\mathbf{q}_f$ is found by,

$$\mathbf{q}_f = \arg\min_{\mathbf{q}} \quad -J(\mathbf{q})$$
$$\text{subject to:} \quad \mathbf{Mq} + \mathbf{m} \leq 0 \tag{7}$$
$$c_{\text{NL}}(\mathbf{q}) \leq 0$$

where $\mathbf{M}$ and $\mathbf{m}$ define the linear constraints and $c_{\text{NL}}(\cdot)$ is a function that defines the nonlinear constraints. By minimizing the negative of the objective functional, the interstory acceleration response is increased and the identification performance should be improved.

Linear Constraints

Two linear constraints are included in the control design. For each closed loop pole of the system $q_i = a_i \pm jb_i$, the following linear constraints are applied.

$$a_i \leq \sigma^c \tag{8a}$$
$$a_i + \zeta^c b_i \leq 0 \tag{8b}$$

where (a_i, b_i) are the real and imaginary components of the pole; $\sigma^c = -5$ is the pole constraint; and $\zeta^c = 10\%$ is the minimum damping ratio for each pole. The first constraint ensures that the closed loop system is stable and achieves a minimum level of settling time for each pole. The second constraint unconservatively ensures that the damping of each mode is greater than ζ^c.

Nonlinear Constraints

To further constrain the control design, nonlinear constraints are used. Three different nonlinear constraints are considered: root mean square (RMS) control force; H_2 constraint; and H_∞ constraint. The first two constraints can be broadly understood as energy constraints and the last constraint is a disturbance rejection constraint. The three constraints are described herein.

The control force constraint ensures that, given an input ground acceleration signal of certain energy, the control force energy will not exceed a certain level. This constraint does not provide an upper bound to the control force commanded, which can result in a saturated control force signal in practice.

TABLE 1: Design results of various observers. The first columns refer to which story-level acceleration sensors are used with zero corresponding to the ground acceleration.

Name	Obs. Vec.	J [·]	$\max_i \sigma_i$ [rad/s]	$\min_i \zeta_i$ [%]	$\|u_c\|_2$ [%]	$\|H\|_2$ [dB]	$\|H\|_\infty$ [dB]
UNC	–	0.92	−0.06	0.34	0.00	37.19	48.78
CT00	full state	1.10	−5.00	9.95	2.00	23.57	13.99
OB00	full accel.	1.10	−5.00	5.60	2.00	23.56	13.99
OB01	[0,3,4]	1.10	−3.68	3.60	2.00	23.56	13.99
OB02	[0,4]	1.10	−2.17	2.14	2.00	23.56	13.99
OB03	[3,4]	1.00	−5.00	9.95	1.77	23.22	14.07
OB04	[4]	0.93	−5.00	8.80	1.70	23.62	14.32

The energy of the control force is determined by finding a solution to the L_2 norm of the control force signal, $\|u_c\|_2$. This is given by multiplying the control force gain matrix by the controllability grammian. Thus,

$$\|u_c\|_2^2 = \mathbf{K}\mathbf{X}_c\mathbf{K}^{\mathsf{T}} \leq f^c$$

where the controllability grammian $\mathbf{X}_c$ is found by solving the Lyapunov equation using the state space matrices of T^c and f^c is the force constraint.

In this study, $\|u_c\|_2$ is computed when the building is subjected to a band-limited white noise ground acceleration signal RMS of 0.1 m/sec² (*i.e.* $\|\ddot{u}_g\|_2 = 0.1$ m/sec²). The control force constraint is 2% of the total weight of the building.

The system energy constraint ensures that the controller does not unduly amplify the output energy. This constraint is understood as a H_2 constraint and requires that the 2-norm of the operator be below a certain threshold. The constraint is given by,

$$\|T^c\|_2 = \sqrt{tr\left\{\mathbf{C}\mathbf{X}_c\mathbf{C}^{\mathsf{T}}\right\}} \leq h_2^c$$

where $h_2^c = 40$ dB is the system energy constraint.

The disturbance rejection constraint ensures that a bounded input will provide a bounded output below a certain threshold. This constraint is understood as a H_∞ constraint applied such that $\|T^c\|_\infty \leq h_\infty^c$. Where, $h_\infty^c = 40$ dB is the constraint, which is implemented by the MATLAB command `norm` for the H_∞ case.

CONTROL WITH DIFFERENT SENSOR CONFIGURATIONS

An observer is designed to estimate the states of the system for feedback control. The observer is characterized by an observer gain matrix **L**. Taken together, the observer and control gain matrices form the feedback control regulator. The resulting system

operator is defined as,

$$T^{\mathrm{o}} := \begin{cases} \begin{pmatrix} \dot{\mathbf{x}} \\ \dot{\hat{\mathbf{x}}} \end{pmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{BK} \\ \mathbf{L\tilde{I}C} & \mathbf{A} - \mathbf{BK} - \mathbf{L\tilde{I}C} \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ \hat{\mathbf{x}} \end{pmatrix} + \begin{bmatrix} \mathbf{G} \\ \mathbf{L\tilde{I}H} \end{bmatrix} \ddot{u}_{\mathrm{g}} + \begin{bmatrix} \mathbf{0} \\ \mathbf{L\tilde{I}} \end{bmatrix} \mathbf{v} \\[12pt] \mathbf{y} = \begin{bmatrix} \mathbf{C} & -\mathbf{DK} \end{bmatrix} \begin{pmatrix} \mathbf{x} \\ \hat{\mathbf{x}} \end{pmatrix} + \mathbf{H}\ddot{u}_{\mathrm{g}} + \mathbf{v} \end{cases} \tag{9}$$

where $(\mathbf{x}^{\mathsf{T}}, \hat{\mathbf{x}}^{\mathsf{T}})^{\mathsf{T}}$ is the augmented state vector including the estimated state vector $\hat{\mathbf{x}}$ and $\tilde{\mathbf{I}}$ contains the rows of an identity matrix corresponding to the measured accelerations defined by $[\tilde{\mathbf{I}}]_{k,l} = \delta_{i_k,l}$ using the Kronecker delta.

The observer is designed using the Kalman filter, which provides the optimal estimate of the state variable (Mendel, 1995). The design is implemented using the MATLAB command `kalman`. This command takes as input the uncontrolled system expressed as a state space model, the input ground acceleration RMS ($\|\ddot{u}_{\mathrm{g}}\|_2 = 0.1$ m/sec^2), and the sensor noise covariance matrix $\mathbf{K_v}$. The output gives the observer gain matrix $\mathbf{L}$. Once the observer gain matrix is designed, the observed system operator T^{o} is constructed as defined in (9).

TABLE 2: Statistics of controlled identification with observers of the second story stiffness presented as percentages.

	Undamaged		Damaged	
	Mean	STD	Mean	STD
UNC	−0.69	12.08	−26.04	5.83
CT00	−0.89	7.61	−25.16	2.11
OB00	−0.70	7.04	−25.18	2.33
OB01	−0.78	7.32	−25.14	2.12
OB02	−1.06	7.95	−25.18	2.37
OB03	−1.43	8.91	−24.96	2.06
OB04	−1.13	8.27	−25.43	3.71

This study will design observers with incomplete sensor measurements. This is accomplished by restricting the measured sensors such that the rows of $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{H}$ are selected corresponding to the measured sensors resulting in $\tilde{\mathbf{C}}$, $\tilde{\mathbf{D}}$, and $\tilde{\mathbf{H}}$, respectively. Then the observer is designed as described.

RESULTS

Using the results of DeVore et al. (2012), controlled substructure identification can be performed on a ROM of the structure. The second story was not experimentally identified by substructure identification so the controller will be designed to improve substructure identification of the second story. Controlled substructure identification performance in undamaged and damaged configurations will be determined using MCS.

The controller is tested in a variety of different sensor configurations. These configurations and the system performance is shown in Table 1. The second story interstory acceleration response is shown in Figure 1. The results show that the different configurations have similar amplification of the second story interstory acceleration.

Once the observers are designed, substructure identification performance is determined for the undamaged and damaged scenarios. The results are summarized in Table 2 and show that the sensor configurations including the ground acceleration are able to provide identification performance comparable to the full state controller. Likewise, sensor configurations excluding the ground acceleration have worse identification performance and see an increase in bias and error variance.

CONCLUSIONS

A controller is developed to improve substructure identification of a the second story, which was not experimentally identified in DeVore et al. (2012). The controller is studied for various sensor arrangements using a Kalman filter to estimate the state vector. For sensor arrangements using measured ground acceleration, substructure identification performance matches (nearly) exactly to the full state feedback controller's performance. For sensor arrangements not utilizing measured ground acceleration, substructure identification performance will vary.

ACKNOWLEDGMENTS

REFERENCES

DeVore, C., Jiang, Z., Johnson, E. A., Christenson, R. E., and Stromquist-Levoir, G. (2012). Experimental verification of substructure identification for damage detection. In *Proceedings of the 2012 Joint Conference of the Engineering Mechanics Institute*, South Bend, IN.

Hou, J., Jankowski, Ł., and Ou, J. (2011). A substructure isolation method for local structural health monitoring. *Str. Control & Health Mon.*, 18(6):601–618.

Koh, C. G., See, L. M., and Balendra, T. (1991). Estimation of structural parameters in time domain: A substructure approach. *Eq. Eng. & Str. Dyn.*, 20(8):787–801.

Mendel, J. M. (1995). *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. Prentice-Hall, Upper Saddle River, NJ.

Xing, Z. and Mita, A. (2012). A substructure approach to local damage detection of shear structure. *Str. Control & Health Mon.*, 19(2):309–318.

Yuen, K. and Katafygiotis, L. (2006). Substructure identification and health monitoring using noisy response measurements only. *Computer-Aided Civil and Infrastructure Engineering*, 21(4):280–291.

Zhang, D. and Johnson, E. A. (2012a). Substructure identification for shear structures: Cross power spectral density method. *Smart Materials and Structures*, 21.

Zhang, D. and Johnson, E. A. (2012b). Substructure identification for shear structures I: Substructure identification method. *Str. Control & Health Mon.*, DOI: 10.1002/stc.1497.

Zhang, D. and Johnson, E. A. (2012c). Substructure identification for shear structures II: Controlled substructure identification. *Str. Control & Health Mon.*, DOI: 10.1002/stc.1498.

Sensor Network for Pavement Performance Monitoring

Ram Rajagopal[1,2], Ronnie Bajwa[2], Erdem Coleri[2], Chris Flores[2], Pravin Varaiya[2]

[1]Stanford Sustainable Systems Lab, Department of Civil and Environmental Engineering, Stanford University, email: ramr@stanford.edu

[2]Sensys Networks; Berkeley, CA email: {rbajwa, ecoleri, cflores, pvaraiya}@sensysnetworks.com

ABSTRACT

Preventive maintenance of roadway pavements requires reliable estimates of pavement deterioration. Existing successful measurements of pavement deterioration rely on knowing the statistical distribution of traffic loads in the roadway. Existing systems are expensive to deploy. This paper describes a sensor network system for road pavement monitoring. It is capable of estimating the load of trucks in real-time and utilizes low deployment and maintenance cost wireless sensors. The approach relies on combining careful pavement models with data measured in real time and provides a reliable estimate of truck loads. Based on this estimates we are able to infer various pavement performance metrics under strict bounds.

INTRODUCTION

Sustainable management of roadway pavements requires substituting costly rehabilitation with appropriate preventive maintenance. Existing preventive maintenance approaches rely on yearly inspections at selected locations and sparse traffic load information to make decisions on when a road section needs intervention. The methodology's performance depends significantly on data availability, which is limited by the high cost of routine inspections and of permanent monitoring systems. Today Caltrans has 70 static weigh stations and 106 weigh-in-motion or WIM stations to monitor California's 50,000 lane miles of highway and 12,488 bridges with a median age of 43 years. An adequate monitoring system for California needs many more weigh stations, but this need cannot be met by today's static or WIM stations, which are expensive to install and maintain.

Instead, we propose SW-WIM (Sensys Wireless WIM), a cost effective embedded wireless sensor network system. The sensor network directly measures the impact of traffic on the road pavement and provides quantitative measures of pavement use. The system consist of accelerometers, magnetometers and temperature sensors embedded flush with the pavement. The sensors report data wirelessly to an access point (AP) installed in the side of road. The system is designed to provide continuous information for five to ten years, depending on the metrics required, relying on batteries. The AP can perform computation, store information and communicate results utilizing either the cellular network or existing network systems. Algorithms in the AP rely on adaptive signal processing techniques to compute important metrics

from the sensor measurements such as average traffic flow, average speed, vehicle traffic classification and accumulated load traveled in the highway. These metrics are then be utilized to determine pavement state and cost impact of this state utilizing existing models for pavement evaluation. Inspection locations can be potentially screened utilizing information from the system. Novel progressive deterioration metrics are also investigated.

There are four areas related to this work: applications of wireless sensor networks (WSN) to transportation, applications of sensor networks in infrastructure monitoring, weigh-in-motion technologies, and algorithms that estimate pavement deflection from acceleration. Applications of WSNs in transportation have been growing. For example, WSNs have been used for vehicle detection using magnetic sensors (Haoui et al., 2008). Much less has been done in terms of monitoring the response of road infrastructure itself. Monitoring large infrastructures using accelerometer sensor networks has been studied for structural monitoring of bridges (Kim et al. 2005; Lynch et al., 2006) and buildings (Lynch et al., 2006). Wired embedded sensors in concrete structures have been investigated (Cebon, 2006) but usually require complex installation procedures and have limited lifetime if used in roads. WIM technologies haven't advanced much in the last decade but there have been attempts to use multiple sensors to filter out dynamic component and not require special-material pavement near the sensor (Cebon, 2006). Estimating pavement deflection (or displacement) from acceleration is a challenging problem in itself. Simple double integration leads to a large unpredictable drift making the resulting signal practically useless (Cebon, 2006).

The rest of the paper is organized as follows. Section 2 describes SW-WIM. Section 3 presents the test results. Section 4 summarizes our conclusions from this study.

SW-WIM

This section provides a description of the SW-WIM system configuration at the I-80 test site, the sensor design, and the signal processing algorithms.

System Configuration

Fig 1 is a schematic of SW-WIM deployed on lane 2 of westbound I-80 in Pinole, CA at the Appian Way exit. This test site is directly downstream of a Caltrans WIM station whose data serves as 'ground truth'. The left side of Fig 1 shows the sensor layout in the 3.66m (12ft) wide lane of the pavement. In the actual deployment each array has a small stagger to prevent 'shadowing' of the pavement vibrations.

There are four arrays of accelerometer sensors (red circles) and four (five) magnetometer sensors (black squares). The arrays are 4.6m (15ft) apart. The multiple array measurements permit separation of the dynamic and static axle load (Cebon, 2006). Current WIMs impose the requirement of smooth pavement to reduce the effect of dynamic component but this is a very expensive approach. The six sensors

in each array are redundant, but allow us to determine the sensitivity of the measurements to sensor placement. Each accelerometer sensor measures the vibration or acceleration of the pavement at the sensor location. The magnetometer sensors measure the arrival and departure times of each vehicle. Since the distance between two magnetometer sensors is fixed, the arrival (or departure times) from any pair of magnetometer sensors gives the vehicle speed.
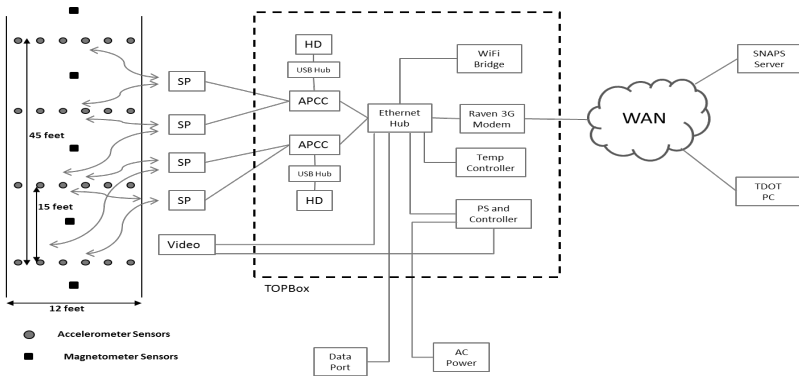
Figure 1 SW-WIM test site on I-80 in Pinole, CA.

The rest of the equipment is mounted on a 5m (15ft) pole on the side of the highway inside a ruggedized 'TOPBox'. This equipment provides remote control and observation of the SW-WIM system. The equipment includes (i) two APCC processor modules with attached Hard Drive (HD) storage that are connected to the external SP radio modules, which receive the sensor data; (ii) external Pan-Tilt-Zoom camera for video capture; (iii) Power Supply (PS) and temperature controller for remote sub-system reset and monitoring; and (iv) cellular 3G modem, Wi-Fi bridge, and Ethernet data port for multiple ways of providing system and data access. The test system is a convenient and versatile platform for collecting pavement vibration data under a variety of traffic and environmental conditions. Data from the site can be collected 24x7.

Sensor Design

Each accelerometer sensor cube is comprised of a MEMS accelerometer, a temperature sensor, a microprocessor, memory, a radio transceiver, an antenna, battery, and an electronic PC board that interconnects these components, assembled as in Fig 2 (left). The assembly is placed in a hardened plastic 7.4 cm x 7.4 cm x 4.9 cm cube (middle), which is then filled with a potting material. The magnetometer sensor cube is virtually identical, except that the accelerometer is replaced by a magnetometer. Each cube is placed in a 5.7-cm diameter, 10-cm deep hole in the pavement (Fig 2, right), which is then filled with fast-drying epoxy. The pavement needs no preparation and a sensor cube is installed in 10 minutes. (The system of Fig

1 was installed by a 4-person crew in four hours.) The analog accelerometer sensor measurements are low-pass-filtered at 50Hz, sampled at 512Hz and sent through a 12-bit A/D converter. In practical applications, it is advised to sample well over Nyquist (at least 5 to 10 times) and in this case additional oversampling also helps us validate the system analysis. The resulting bit stream is transmitted by the sensor radio to the APCC. The APCC has a Linux processor and various communications interfaces as shown in Fig 1. The Data from the APCC and the camera can be stored locally and downloaded over the 3G network.
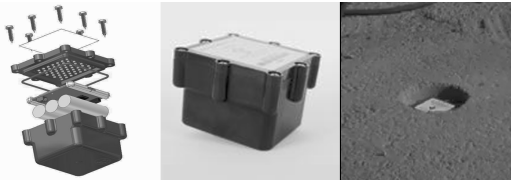
Figure 2 Sensor cube assembly (left), package (middle), and installation (right).

Sensor durability is enhanced by a protective casing and internal potting designed to endure three to four times the maximum legal truck loads. Power consumption is minimized by carefully programming the sensor network to utilize a TDMA wireless communication protocol [see 2 for details]. Sensor information is transmitted in slotted time units, so each sensor wakes up at a pre-determined time, and transmits its information alone preventing wireless collisions. All sensors synchronize clocks every second by acquiring a beacon from the access point. In raw data mode, the sensor lifetime is observed to be 2 years for the magnetometers and about a year for the accelerometers. If sensors are programmed to compute calculations before transmission, it has been shown the magnetometers last 10 years (Haoui et al., 2008) and accelerometers last 4 years.

The APCC processes the data from the sensor networks to determine load estimates. The algorithms rely on various properties of the pavement and are described next.

Axle Detection and Speed Estimation

Speed is estimated by utilizing the difference in detection time of a vehicle by two magnetometers separated by a known distance (Haoui et al., 2008). Axle counts are estimated utilizing signal energy peak detection algorithms on the first array of accelerometers (Bajwa et al., 2006). Both algorithms are implemented utilizing simple local filtering operations, which significantly reduce the amount of data that needs to be transmitted to the APCC for each truck. Usually it reduces to two samples for each magnetometer and about six samples for the axle detection.

Load Estimation

The simplest model for pavement-vehicle interaction models the pavement as a composite one dimensional Euler beam resting on a Winkler Foundation (Cebon,

2006), and each axle of the vehicle is a moving force modulated by its suspension system. Road roughness interacting with the vehicle suspension causes the applied load by the vehicle to be given by $F(t) = F_s + F_d \cos(\omega t)$, where F_s is the static load, F_d the dynamic load (about 30% of F_s) and ω the suspension frequency. The vertical displacement response the pavement is then of the form $y(t) = F\Phi(vt)$, and $\Phi(x)$ is strongly dependent on the structural and material properties of the pavement (Rajagopal, 2009). The typical response is shown in Figure 3. Based on the theory developed in (Rajagopal, 2009), the response can be approximated utilizing a Gaussian PDF, so $\Phi(x) = K\exp(-x^2/2\sigma^2)$ for an appropriate constant K. It can be shown that the second derivative of $y(t)$ is given by $a(t) = \alpha\psi(t,\sigma)$ and $F = \beta\dfrac{\alpha}{v^2}$, where β is a constant that does not depend on speed, and only depends on the pavement.
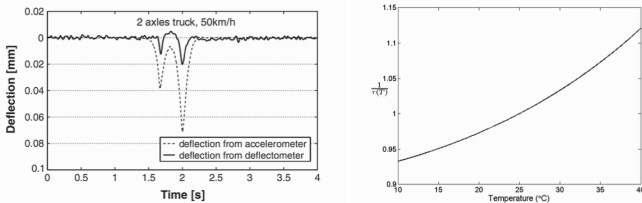
Figure 3: (a) Typical pavement response of a truck and (b) LET model estimate of temperature correction factor.

The load to pavement transfer function depends on temperature. Given the material and the design thickness of the pavement at a road site, a layered elastic theory model (LET) was developed to predict the temperature response. The theory predicts that the deflection is affected multiplicatively by temperature. The temperature factor $\tau(T)$ relative to the response at 25°C can be found for the site as shown in Figure 3.

The complete relationship between load, speed, pavement parameters and peak acceleration estimate is given by

$$F = \beta\frac{\alpha}{v^2}\tau(T)$$

The parameters for the model need to be estimated from calibration data. Later we show experimentally that the error between estimated load $\hat{F}$ and actual load F is given by $F = (1+\varepsilon[\%]/100)\hat{F}$, where $\varepsilon[\%]$ is a normally distributed relative error with variance $10^4\sigma_k^2$ (which is normalized so the variance constant is in relative units).

The actual load estimation algorithm needs to work with noisy acceleration measurements as shown in Figure 5(a), top. In order for the algorithm to be applied in actual data, we perform three steps: (a) filter each sensor's acceleration measurement using a low pass filter with bandwidth 50Hz, (b) align the signal responses from

different sensors, since the truck reaches each array at a different time and (c) average the responses of all sensors. This results in a single average acceleration profile $a_m(t)$ shown in Figure 5(a) bottom. The last step of the algorithm consists in fitting a *multiple axle model* to the data by minimizing the mean squared error:

$$(\alpha_i^*, \sigma_i^*, \mu_i^*) = \arg \min_{\alpha_i, \sigma_i, \mu_i} \int_{-\infty}^{\infty} \left(a_m(t) - \sum_{i=1}^{3} \alpha_i \Psi(t - \mu_i, \sigma_i) \right)^2 dt$$

The equation above applies for a 3-axle truck, but more axles are adjusted accordingly. An important observation is that the above problem is not convex, but initializing the algorithm with axle locations from the axle detection step results in very good fits as evidenced by Figure 5.

Pavement Damage Indicators

The most well known pavement damage indicator is the Present Serviceability Index (Cebon, 2006) or PSI. It is a number between 0 and 5 that determines the state of the road and can be directly related to cracking and patching, rut depth and surface roughness of the road via a regression equation. The decrease in PSI is related to the Equivalent Standard Axle Loads (ESAL). The ESAL captures the damage effect of an arbitrary (heavy) axle load by transforming it into the number N of equivalent 18,000 lbs single axle loads, given by:

$$N = \left(\frac{F}{80kN} \right)^4$$

The total number N accumulated over all trucks passing the site gives the site ESAL which can be used to determine damage state of the roadway using different models according to the AASHTO manual. SW-WIM provides an estimate $\hat{F}$ of the load F of each axle with a multiplicative error model described in the previous subsection. The EASL is then a random variable given by

$$\hat{N}_i = \left(\frac{\hat{F}_i}{80\text{kN}} \right)^4 = (1 + k_i)^4 \left(\frac{F_i}{80\text{kN}} \right)^4 \approx (1 + 4k_i + 6k_i^2)N_i$$

and $k = (1 + \varepsilon[\%]/100)$ with the approximate very good for the range [-0.3,0.3]. If we interpret each truck EASL as arising from a distribution of EASL values, and that multiplicative errors are independent of load values, we obtain that for large number B of trucks:

$$\frac{1}{B} \sum_{i=1}^{B} \hat{N}_i \approx (1 + 6\sigma_k^2) \frac{1}{B} \sum_{i=1}^{B} N_i \Rightarrow \sum_{i=1}^{B} N_i = \sum_{i=1}^{B} \frac{\hat{N}_i}{(1 + 6\sigma_k^2)}$$

Thus a simple bias correction of the EASL gives an improved estimate of total EASL. Notice that from the system performance statistics in the load estimation we have a good estimate of the variance in the denominator.

EXPERIMENTAL RESULTS
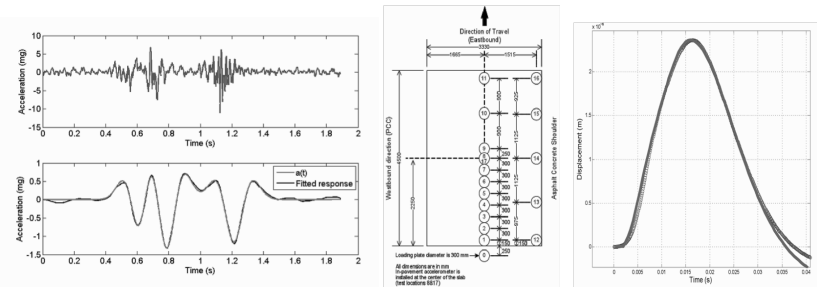
Figure 5: (a) Raw acceleration signal and averaged acceleration response for all arrays and fitted response (b) Measurement plan and displacement estimates for a Falling Weight Deflectometer experiment.

In an Falling Weight Deflectometer (FWD) test, weights are dropped at a location in the road, and displacements estimated from deflectometers on top of the road. Fig. 5(b) compares displacement estimates from our system and the FWD data, showing excellent agreement.

	Temperature compensation		No compensation	
	Mean Error (%)	Std of errors (%)	Mean Error (%)	Std of errors (%)
Axle 1	-0.27	5.25	-0.44	6.66
Axle 2	-0.22	4.75	-0.39	6.27
Axle 3	-0.33	5.81	-0.43	6.63
Total	-0.19	4.09	-0.33	5.61

	Wireless WIM error	Conventional WIM error	Maximum allowed error
Axle 1	11.29	18.67	20
Axle 2	10.07	26.49	15
Axle 3	12.44	37.35	15
Total	8.76	23.23	10

Figure 6: (a) Performance statistics (in %) for each axle in different scenarios **(b)** LTTP error estimate which equals the absolute mean error plus 1.96 times standard deviation.

More detailed analysis was performed by collecting ground truth data from a static weight station for 52 3-axle trucks on the site, and evaluating the performance of the load estimation procedure. The results were also compared to a standard WIM station. Both are shown in Figure 6. Temperature compensation is definitely required for good performance, and works surprisingly well with the nominal road pavement design parameters for the site. Notice that SW-WIM meets the WIM standard requirements for the state of California, but the standard WIM does not.

The std. errors in Figure 6(a) can be used to correct EASL estimates. Figure 7 shows the ground truth and estimates, and a smoothed histogram of the error distributions. Notice they are approximately normal for axles 2 and 3, which are 'heavy axles'.
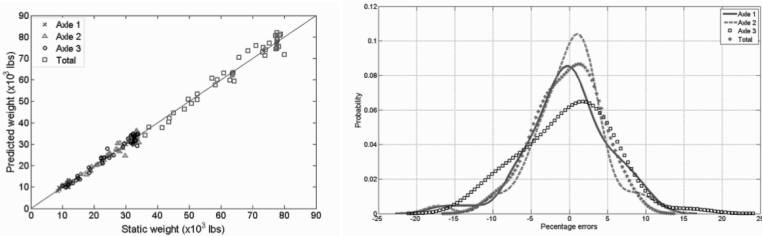
Figure 7. (a) Ground truth vs. estimate for individual axles and total truck load for 52 trucks. **(b) Error histograms showing approximate normality for Axles 2 and 3 (heavy axles).**

CONCLUSION

The proposed system accurately estimates the individual axle loads of trucks, and enables accurate estimation of the total EASL. More interestingly, direct estimate of the pavement parameters is enabled by the system using a novel regression methodology. In future work, we propose investigating the dynamic load characteristics and utilizing them to predict rutting and roughness directly, instead of utilizing EASL estimates. We will also test in-sensor estimation of loads to improve lifetime of the accelerometer sensors.

REFERENCES
1. Cebon, D. (2006), *Handbook of Vehicle-Road Interaction*, Taylor and Francis.

2. Haoui, A., Kavaler, R., and Varaiya, P. (2008), "Wireless magnetic sensors for traffic surveillance", *Transportation Research Part C*, 16, 294–306.

3. Bajwa, R., Rajagopal, R., Varaiya, P., and Kavaler, R. (2010), "In-pavement wireless sensor network for vehicle classifcation", *10th International Conference on Information Processing in Sensor Networks* (IPSN2010), Chicago, Illinois, USA, April 12 - 14, 2011.

4. S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, and M. Turon. Health monitoring of civil infrastructures using wireless sensor networks. *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*, pages 254–263, New York, NY, USA, 2007. ACM.

5. Lynch, J.P. and Loh, K.J., A summary review of wireless sensors and sensor networks for structural health monitoring, *Shock and Vibration Digest*, Vol. 38, No. 2, pgs. 91—130, 2006.

6. Rajagopal, R. *Large monitoring systems: data analysis, design and deployment*, Ph.D. dissertation, UC Berkeley EECS, 2009.

Modeling of nonlinear guided waves and applications to structural health monitoring

Claudio Nucera[1] and Francesco Lanza di Scalea[2]

[1]University of California, San Diego, Department of Structural Engineering, 9500 Gilman Drive, M.C. 0085, La Jolla, California, 92093-0085 USA; email: cnucera@ucsd.edu
[2]University of California, San Diego, Department of Structural Engineering, 9500 Gilman Drive, M.C. 0085, La Jolla, California, 92093-0085 USA; email: flanza@ucsd.edu

ABSTRACT

Propagation of nonlinear waves in solid waveguides is a branch of wave mechanics that has received an ever increasing interest in the last few decades. Nonlinear guided waves are promising candidates for interrogating long waveguide-like structures as they conveniently combine high sensitivity to peculiar structural conditions (defects, quasi-static loads, instability conditions), typical of nonlinear parameters, with large inspection ranges, characteristic of wave propagation in confined media. However, the mathematical framework governing the nonlinear guided wave phenomena becomes extremely difficult when characterized to waveguides that are complex in either material (damping, anisotropy, heterogeneous, and the like) or geometry (multilayers, geometric periodicity, and the like). Therefore the successful use of nonlinear guided waves as a structural diagnostic tool is not always a trivial task. In particular, the efficiency of nonlinear Structural Health Monitoring (SHM) techniques based on higher-harmonics generation (harmonics generated by a monochromatic input in nonlinear waveguides) strongly relies on the correct identification of favorable combinations of primary and resonant double-harmonic nonlinear wave modes for which the nonlinear response is cumulative. The present work develops predictions of nonlinear second-harmonic generation and identifies these combinations of wave modes in complex waveguides by extending the classical Semi-Analytical Finite Element formulation to the nonlinear regime, and implementing it into a highly flexible commercial Finite Element code. The proposed algorithm is benchmarked for the following case-studies: a railroad track, a composite quasi-isotropic laminate, and a reinforced concrete slab.

INTRODUCTION

Traditionally the structural monitoring via ultrasounds has been accomplished measuring "linear" parameters of the waves (amplitude, speed, phase shifts) to infer salient features of the inspected structure. However, it is well documented (Dace *et al.*, 1991) that "nonlinear" parameters are, in general, much more sensitive to structural conditions than their linear counterparts. Furthermore the use of nonlinear guided waves is extremely attractive because guided waves combine the mentioned high sensitivity typical of nonlinear parameters with large inspection ranges (Cawley and Alleyne, 1996; Rose, 2002; Bermes *et al.*, 2007). The aforementioned complexity of the mathematical framework governing nonlinear guided wave propagation limited

most of the previous works on elastic waves to the linear elastic regime with the assumption of infinitesimal deformations. However, different mechanisms (increase in wave amplitude, finite deformations, nonlinear strain energy potentials) can give rise to nonlinear effects that eventually become of primary importance. Hence the governing equations need to be modified accordingly: cubic (and possibly higher-order) terms must be included in the elastic energy density expression (Goldberg, 1960; Zarembo and Krasil'nikov, 1971). Among the manifestations of the nonlinear behavior, higher harmonic generation is considered in particular. In this scenario, an initially sinusoidal stress wave of a given frequency distorts as it propagates, and energy is transferred from the fundamental frequency, ω, to the higher harmonics, 2ω, 3ω and so on. For a practical use, this nonlinearity can be quantified via an ultrasonic nonlinear parameter, β, well documented in literature (Bermes *et al.*, 2007). While several investigations pertaining to nonlinear effect in solids and second harmonic generation were reported in the past (de Lima and Hamilton, 2003; Deng, 2003), most of them were limited in their applicability to structures with simple geometries (plates, rods, shells) where analytical solutions for the primary (linear) wave field are available in literature. In the present work the propagation of waves in nonlinear solid waveguides with complex geometrical and material properties is investigated theoretically and numerically. For the solution of the nonlinear boundary value problem, perturbation theory and modal expansion are used (de Lima and Hamilton, 2003). A numerical algorithm is introduced in order to efficiently predict and explore the nonlinear wave propagation phenomena in structural waveguides of different complexity. After a brief description of the theoretical background, three case-studies have been analyzed, namely a railroad track, a composite quasi-isotropic laminate, and a reinforced concrete slab. Favorable combinations of primary and resonant secondary modes (nonlinear resonance conditions) were successfully identified for these complex waveguides. The knowledge of these nonlinear resonance conditions is of primary importance for the actual implementation of structural diagnostic systems based on nonlinear ultrasonic guided waves features.

WAVES IN NONLINEAR ELASTIC REGIME – INTERNAL RESONANCE

In presence of finite deformations, large amplitude waves, nonlinear strain energy potentials and similar nonlinear mechanisms, the generalized Hooke's Law no longer applies and must be replaced by a proper nonlinear constitutive law. Assuming that the body is homogeneous, isotropic and hyperelastic, it possesses a strain energy density ε that is an analytic function of the Green-Lagrange Strain Tensor E such that the Second Piola-Kirchoff Stress Tensor S can be expressed as:

$$S_{ij} = \rho_0 \frac{\partial \varepsilon}{\partial E_{ij}} \tag{1}$$

where ρ_0 is the initial density. It is known that in this scenario the strain energy can be expressed as:

$$\varepsilon = \frac{1}{2}\lambda I_1^2 + \mu I_2 + \frac{1}{3}C I_1^3 + B I_1 I_2 + \frac{1}{3} A I_3 + O\left(E_{ij}^4\right) \tag{2}$$

where $I_1 = E_{ii}$, $I_2 = E_{ij}E_{ji}$, $I_3 = E_{ij}E_{jk}E_{ki}$, λ and μ are the Lamé elastic constants and A, B and C are the Landau-Lifshitz third-order elastic constants (Landau and Lifshitz, 1959). In Eq. (2) first order material nonlinearity was introduced. By substituting Eq.

(2) into Eq. (1) and keeping up to second-order terms in E_{ij} we obtain the nonlinear stress-strain relation

$$S_{ij} = \lambda E_{kk}\delta_{ij} + 2\mu E_{ij} + \delta_{ij}\left(CE_{kk}E_{ll} + BE_{kl}E_{lk}\right) + 2BE_{kk}E_{ij} + AE_{jk}E_{ki} \qquad (3)$$

Using Eq. (3) inside the general momentum equation the Nonlinear Boundary Value Problem governing the propagation of nonlinear elastic waves in isotropic, homogeneous and hyperelastic waveguides can be formulated in vector form as:

$$(\lambda + 2\mu)\nabla(\nabla \cdot u) - \mu\nabla\times(\nabla\times u) + f = \rho_0 \frac{\partial^2 u}{\partial t^2} \qquad (4)$$

$$S^L(u) \cdot n_r = -\overline{S}(u) \cdot n_r \quad on \quad \Gamma \qquad (5)$$

where u is the particle displacement vector, ρ_0, λ and μ are the defined above, f is the nonlinear term acting as a body force, n_r is the unit vector normal to the surface of the waveguide Γ and S_L and $\overline{S}$ are the linear and nonlinear parts of the second Piola-Kirchoff stress tensor, respectively. The nonlinear boundary value problem in Eqs. (4)-(5) is solved in the following using perturbation theory, which is based on writing the solution as sum of two terms, namely $u=u^{(1)}+u^{(2)}$, where $u^{(1)}$ is the primary solution and $u^{(2)}$ is the secondary solution due to nonlinearity, this assumed to be small compared to $u^{(1)}$ (*perturbation condition*). Using this condition the original nonlinear boundary value problem is divided into two linear boundary value problems, namely the first-order and second-order approximations. Following (Auld, 1990) and (de Lima and Hamilton, 2003), if ω is the primary frequency that we suppose to convey into the system via a monochromatic wave, the first order nonlinear solution is calculated through modal expansion using the existing propagating guided modes 2ω as:

$$v(x, y, z, t) = \frac{1}{2}\sum_{m=1}^{\infty} A_m(z)v_m(x, y)e^{-i2\omega t} + c.c. \qquad (6)$$

where $c.c.$ denotes complex conjugates, v_m is the particle velocity vector referred to the mth mode at 2ω and A_m is the higher order modal amplitude given by

$$A_m(z) = \overline{A}_m(z)e^{i(2kz)} - \overline{A}_m(0)e^{ik_n^* z} \qquad (7)$$

All the details regarding the calculation of the modal amplitudes can be found in (de Lima and Hamilton, 2003). At this stage it is important to emphasize how the internal resonance mechanism, which in turn produces a cumulative nonlinear response, relies on the simultaneous occurrence of two conditions, namely:

1. Phase Matching: $k_n^* = 2k$
2. Non-zero power transfer from primary to secondary wave: $f_n^{surf} + f_n^{vol} \neq 0$

CO.NO.SAFE ALGORITHM

The Semi-Analytical Finite Element formulation (*S.A.F.E.*) in its linear fashion has been extensively discussed in the past highlighting its great potential in efficiently calculating the dispersion properties of waveguide-like structures. These properties are crucial for the implementation of any SHM system based on the use of ultrasonic guided waves. Concerning the detail of the classical *SAFE* formulation, the interested reader is referred to (Bartoli *et al.*, 2006). In the present work the classical linear

formulation is extended to the nonlinear regime according to the theory discussed in the previous section. The resultant **No**nlinear **S.A.F.E.** formulation is implemented into **CO**MSOL commercial code (***CO.NO.SAFE***). In this way the full power of existing libraries and routines of the commercial code is exploited and the internal resonance conditions of several structural waveguides of different complexity can be studied in a straightforward manner via user-friendly interfaces. Furthermore, since all the nonlinear parameters involve gradients of the displacement field up to the third order (de Lima and Hamilton, 2003), high-order finite elements (at least cubic) need to be used in order to obtain meaningful results; this implement is not always trivial for general non-commercial SAFE algorithm. Starting from the nonlinear boundary value problem described in Eqs. (4)-(5), the displacement field is approximated in the cross-section of the waveguide and is enforced to be harmonic in time and along the direction of wave propagation. For the generic *e*th element this reads:

$$u^e = N^e U^e e^{i(kz - \omega t))} \tag{8}$$

where $N^e = N^e(x,y)$ is the matrix of shape functions and U^e is the nodal displacement vector for the *e*th element. Assuming this displacement field in (4-5) is the only trivial modification that needs to be done in the original FEM formulation. The original quadratic eigenvalue problem is linearized doubling the space dimension (Bartoli *et al.*, 2006). The nonlinear boundary value problem has been implemented in COMSOL using the general PDE solver engine. The COMSOL formalism for the boundary value problem with Neumann B.C. (which corresponds to the guided wave propagation since a stress-free B.C. needs to be applied) is:

$$\nabla \cdot \left(c\nabla U + \alpha U - \gamma \right) - \beta \cdot \nabla U - aU + \lambda d_a U = 0 \tag{9}$$

$$n \cdot \left(c\nabla U + \alpha U \right) + qU = 0 \tag{10}$$

where U represents the set of dependent variables to be determined and all the remaining quantities are matrix coefficients admitting complex values, which is essential for viscoelastic materials. The nonlinear part of the algorithm has been coded in MATLAB and connected to COMSOL using the *LiveLink* package. In this way, once the dispersion properties are obtained, several possible combinations of potential resonant modes are selected and analyzed. Resonant modes are identified making use of the phase-matching and non–zero power transfer conditions discussed before. The possibility to assess internal resonance conditions for complex solid waveguides is crucial for the efficiency of a given SHM inspection approach. In fact, once a resonant combination of modes is identified, nonlinear response becomes cumulative, namely the amplitude of the second harmonic grows with distance (de Lima and Hamilton, 2003). In this way the efficiency of the given nonlinear technique is dramatically maximized.

APPLICATIONS
1. Railroad track

The widely used 136-lb RE railroad track was considered as first case-study. Due to the complex geometry of the cross section, dispersion curves and, consequently, higher harmonic generation conditions cannot be calculated analytically. The material properties considered are given in Table 1. The Landau-Lifshitz third-order elastic constants are detailed in (Sekoyan and Eremeev, 1966).

Table 1. Material properties assumed for the railroad track analysis.

ρ [kg/m^3]	λ [GPa]	μ [GPa]	A [GPa]	B [GPa]	C [GPa]
7932	116.25	82.754	-340	-646.667	-16.667

The complexity of the guided wave propagation for this particular waveguide is evident considering the abundance of possible propagative modes present (Fig. 1a) and the complexity of the phase-velocity dispersion curves (Fig. 1b), especially at higher frequencies.
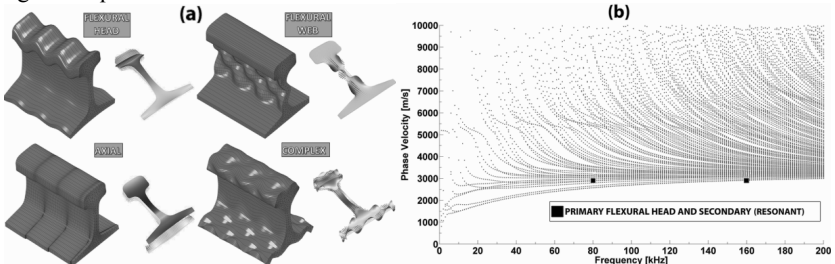
Fig. 1. (a) Exemplary propagative modes in the rail. Phase-velocity dispersion curves in the (0-200) kHz frequency range with primary and secondary modes highlighted.

A flexural vertical mode was selected as primary excitation at 80 kHz. The results of the CO.NO.SAFE analysis disclosed the presence of some synchronous secondary modes at 160 kHz with one in particular (slightly different flexural vertical type) able to produce internal resonance. Fig. 2a and Fig. 2b display the selected modes, while Fig. 2c spotlights the very high value of modal amplitude related to the only secondary resonant mode; small amplitude values associated to the other synchronous modes, for which power transfer is absent, are also shown in the same figure. The previous results point up a favorable combination of primary and secondary wave fields able to maximize the nonlinear response of the waveguide.
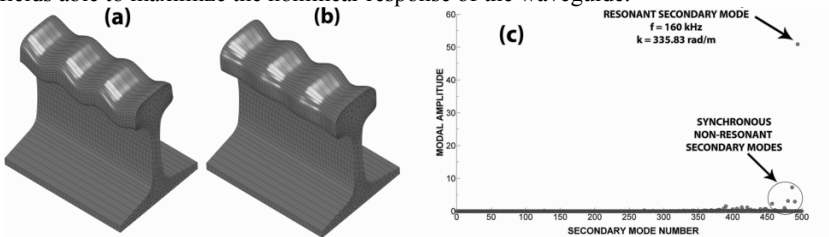
Fig. 2. (a) Selected primary mode at 80 kHz. (b) Resonant secondary mode at 160 kHz. (c) Modal amplitude plot for secondary propagative modes.

2. Anisotropic Elastic Composite Laminate

A multi-layered composite laminate with unidirectional laminae in a quasi-isotropic layup was examined next. More specifically, the selected system consists of eight unidirectional T800/924 graphite-epoxy plies with a stacking sequence of $[\pm45/0/90]_S$. Each layer has a thickness of 0.125 mm resulting in a total laminate

thickness of 1 mm. The material properties for each single lamina in the principal directions of material symmetry are: $\rho = 1500$ kg/m^3, $E_{11} = 161$ GPa, $E_{22} = 9.25$ GPa, $G_{12} = 6.0$ GPa, $v_{12} = 0.34$ and $v_{23} = 0.41$ (Percival and Birt, 1997). The stiffness matrix for each lamina was rotated according to the angle between the fiber direction and the wave propagation direction (Bartoli *et al.*, 2006). In the following, wave propagation was assumed at 0° with respect to the fiber direction 1 (the extension to cases where this angle assumes different values is trivial). After all the matrices were rotated, the governing eigenvalue problem was solved as in the previous sections using the rotated stiffness matrices in the constitutive relations. Periodic Boundary Conditions (PBCs) were employed to gain computational efficiency (Predoi *et al.*, 2007). The third-order elastic constants assumed for each lamina are: $A = 15$ GPa, $B = -33$ GPa and $C = -14$ GPa (Prosser, 1987). The nonlinear post-processing was developed between 2.5 MHz and 5.0 MHz. A complex primary mode combining attributes typical of axial and flexural horizontal modes was selected as input. One of the propagative modes at the double harmonic (5 MHz) was found able to produce internal resonance. The results, in terms of modal amplitude plots, are shown in Fig. 3 along with the primary and secondary modeshapes. It can be noted from Fig. 3 how drastic is the predominance of the only resonant mode in terms of modal amplitude, when compared to all the other propagative secondary modes existing at 5 MHz. Both primary and secondary modes concentrate the wave energy near the center of the waveguide; consequently, this combination appears appealing for the inspection of the laminate because of the expected reduced wave leakage into surrounding areas.
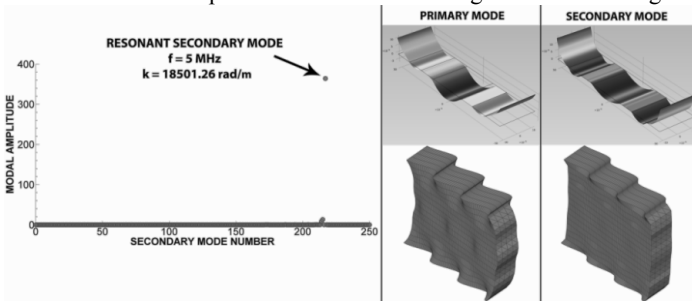
Fig. 3. Modal amplitude plot for secondary propagative modes, along with contour plots and 3D views of the selected primary and secondary modes for the elastic composite laminate.

3. *Reinforced Concrete Slab*

The complexity here arises from the coexistence of two domains with very different material properties. Previous studies have shown the influence of the reinforcement on the dispersion curves (Predoi *et al.*, 2007). The present work analyzes the nonlinear features of the guided wave propagation for this particular heterogeneous system. Likewise the previous case, PBCs are used to model the geometrical periodicity. The 2D periodic cell considered is 6 cm wide and 8 cm tall. The steel bars are assumed to be 1.6 cm in diameter (Fig. 4a). Material properties assumed for the concrete domain are: $\rho = 2133$ kg/m^3, $C_{11} = 33.2$ GPa, $C_{66} = 11.8$ GPa (Bouhadjera, 2004). For the steel bars, the following values were used: $\rho = 7900$

kg/m^3, C_{11} = 280 GPa, C_{66} = 80 GPa (Predoi *et al.*, 2007). The CO.NO.SAFE algorithm was used with 40 kHz as the primary frequency. The primary mode selected as input exhibits essentially a flexural horizontal displacement field.
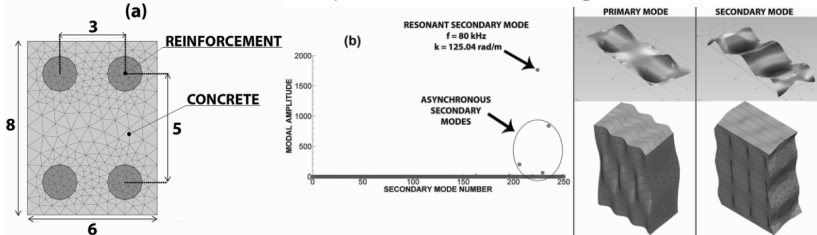
Fig. 4. (a) Geometrical details and finite element mesh for the periodic cell representative of a 8 cm thick reinforced concrete slab (dimensions in cm). (b) Modal amplitude plot for secondary propagative modes along with contour plots and 3D views of the selected primary and secondary modes for the reinforced concrete slab.

The nonlinear results are presented in Fig. 4. They reveal the presence of few asynchronous modes characterized by relatively large power transfer (modal amplitude values inside the circle) and only a single resonant secondary mode able to verify also the phase-matching condition. The nature of this identified advantageous combination of modes is represented in Fig. 4.

CONCLUSIONS

The use of nonlinear guided waves is gaining increasing attention in the non-destructive evaluation and structural health monitoring communities. Proper application of nonlinear measurements requires a thorough understanding of the higher-harmonic generation phenomena that can be expected for the test waveguide. In the present work, the classical S.A.F.E. algorithm was extended to the nonlinear regime and implemented in a powerful multipurpose commercial FEM code (COMSOL). The result is a new tool that opens new possibilities for the analysis of dispersion characteristics and, most importantly here, nonlinear internal resonance conditions, for a variety of complex structural waveguides that do not lend themselves to alternative analyses such as purely analytical solutions.

The specific "complex" cases that were examined include: complex geometry (railroad track), multilayered composite panels (8-ply quasi-isotropic laminate), and heterogeneous periodic systems (reinforced concrete slab). In all these cases, the proposed algorithm successfully identified appropriate combinations of resonant primary and secondary modes that exhibit the desired conditions of synchronicity and large cross-energy transfer. These properties can be exploited in an actual system aimed at monitoring the structural condition of the waveguide by nonlinear waves (detect defects, measure quasi-static loads or instability conditions, etc.).

REFERENCES

Auld, B. A. (**1990**). *Acoustic fields and waves in solids* (R.E. Krieger, Malabar, Fla.).

Bartoli, I., Marzani, A., di Scalea, F. L., and Viola, E. (**2006**). "Modeling wave propagation in damped waveguides of arbitrary cross-section," J Sound Vib **295**, 685-707.

Bermes, C., Kim, J. Y., Qu, J. M., and Jacobs, L. J. (**2007**). "Experimental characterization of material nonlinearity using Lamb waves," Appl Phys Lett **90**.

Bouhadjera, A. (**2004**). "Simulation of In-Situ Concrete Conditions Using a Novel Ultrasonic Technique," Proceedings of 16th World Conference on Non-Destructive Testing.

Cawley, P., and Alleyne, D. (**1996**). "The use of Lamb waves for the long range inspection of large structures," Ultrasonics **34**, 287-290.

Dace, G., Thompson, R., Rehbein, D., and Buck, O. (**1991**). "Nonlinear acoustic, a technique to determine microstructural changes in material," Rev. Prog. Quant. NDE **10B**, 1685-1692.

de Lima, W. J. N., and Hamilton, M. F. (**2003**). "Finite-amplitude waves in isotropic elastic plates," J Sound Vib **265**, 819-839.

Deng, M. X. (**2003**). "Analysis of second-harmonic generation of Lamb modes using a modal analysis approach," J Appl Phys **94**, 4152-4159.

Goldberg, Z. A. (**1960**). "Interaction of plane longitudinal and transverse elastic waves," Sov. Phys. Acoust., 306-310.

Landau, L. D., and Lifshitz, E. M. (**1959**). *Theory of elasticity* (Addison-Wesley Pub. Co., London).

Percival, W. J., and Birt, E. A. (**1997**). "A study of Lamb wave propagation in carbon-fibre composites," Insight **39**, 728-735.

Predoi, M. V., Castaings, M., Hosten, B., and Bacon, C. (**2007**). "Wave propagation along transversely periodic structures," J Acoust Soc Am **121**, 1935-1944.

Prosser, W. H. (**1987**). "Ultrasonic Characterization of the Nonlinear Elastic Properties of Unidirectional Graphite/Epoxy Composites," NASA Contractor Report **4100**, 75-120.

Rose, J. L. (**2002**). "Standing on the shoulders of giants: An example of guided wave inspection," Mater Eval **60**, 53-59.

Sekoyan, S. S., and Eremeev, A. E. (**1966**). "Measurement of the third-order elasticity constants for steel by the ultrasonic method," Measurement Techniques **0543-1972**, 888-893.

Zarembo, L. K., and Krasil'nikov, V. A. (**1971**). "Nonlinear phenomena in the propagation of elastic waves in solids," Soviet Physics USPEKHI **13**, 778-797.

Multivariate Analysis and Prediction of Wind Turbine Response to Varying Wind Field Characteristics Based on Machine Learning

J. Park[1], K. Smarsly[1], K. H. Law[1] and D. Hartmann[2]

[1]Dept. of Civil and Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA; email: {jkpark11, smarsly, law}@stanford.edu
[2]Dept. of Civil and Environmental Engineering, Ruhr University Bochum, Universitätsstr. 150, 44780 Bochum, GERMANY; email: hartus@inf.bi.rub.de

ABSTRACT

Site-specific wind field characteristics have a significant impact on the structural response and the lifespan of wind turbines. This paper presents a machine learning approach towards analyzing and predicting the response of wind turbine structures to varying wind field characteristics. Machine learning algorithms are applied (i) to better understand changes of wind field characteristics due to atmospheric conditions, and (ii) to gain new insights into the wind turbine loads being affected by fluctuating wind. Using Gaussian Mixture Models, the variations in wind field characteristics are investigated by comparing the joint probability distribution functions of several wind field features, which are constructed from long-term monitoring data taken from a 500 kW wind turbine in Germany. Furthermore, based on Gaussian Discriminative Analysis, representative daytime and nocturnal wind turbine loads are predicted, compared, and analyzed.

INTRODUCTION

Variations in wind field characteristics can affect the structural and operational response of wind turbines. Wind field characteristics are often described by time-dependent statistical parameters such as mean wind speed, turbulence intensity, mean wind direction and vertical mean wind profile, which depend on the surface roughness (e.g. land or water) and on the atmospheric stability (e.g. day or night). For example, the daytime and the nocturnal wind field characteristics can differ significantly. Knowledge on the fluctuating wind field and its effects on wind turbines are essential for designing as well as for cost-efficiently managing wind turbines. However, these effects have rarely been studied.

This paper presents a machine learning approach for analyzing and predicting the response of wind turbine structures to varying wind field characteristics. First, an integrated life-cycle management (LCM) framework, which provides long-term monitoring data recorded from a wind turbine, is briefly introduced. The proposed machine learning approach to relate wind field effects on wind turbines is discussed in details. Using the monitoring data provided by the LCM framework, a case study

investigating the variation between the daytime and nocturnal wind field characteristics is presented.

WIND TURBINE LIFE-CYCLE MANAGEMENT FRAMEWORK

The wind turbine LCM framework, installed on the aforementioned 500 kW wind turbine, consists of (i) a structural health monitoring (SHM) system; and (ii) several software modules installed at spatially distributed locations. The software modules include a monitoring database for persistent storage of the recorded data sets, a central server for automated data processing, a management module supporting life-cycle analyses, and Internet-enabled user interfaces providing online access to authorized users and to external application programs. Details on the LCM framework and on the integrated software modules have been described in Smarsly *et al.* (2011a,b, 2012a,b,c) and Hartmann *et al.* (2011). In the following subsections, the SHM system is introduced, followed by a description of the data sets used in this study.

Structural Health Monitoring System. The structural health monitoring system comprises of a network of sensors, data acquisition units and an on-site server located in the wind turbine. The wind turbine has a hub height of 65 m and a rotor diameter of 40.3 m. The sensors (accelerometers, displacement transducers, and temperature sensors) are placed at different levels on the inside and the outside of the steel tower and on the foundation of the wind turbine (Figure 1). In addition, two anemometers are deployed to continuously measure the wind speed, the wind direction, and the air temperature. The first anemometer, a cup anemometer, is installed on top of the wind turbine nacelle (h = 67 m). Additionally, the second anemometer, a three-dimensional ultrasonic anemometer, is mounted on a telescopic mast adjacent to the wind turbine (h = 13 m).

The data acquisition units, installed in the maintenance room of the wind turbine, continuously collect, sample, and digitize the structural and environmental data. Referred to as "primary monitoring data", the data sets are continuously forwarded from the data acquisition units to the on-site server for temporary storage and periodic local backups. Through a permanently installed DSL connection, the on-site server transfers the primary monitoring data to the monitoring database. The database, which is an integral part of the LCM framework, is installed at the Institute for Computational Engineering (ICE) at the Ruhr University Bochum. In addition to the structural and environmental data, operational data is also transferred to the monitoring database. The operational data, such as power production of the wind turbine as well as revolutions and pitch angles of the rotor, are recorded by the wind turbine machine control system. Once being stored in the database, the monitoring data are remotely available online and accessible by authorized personnel and software modules.
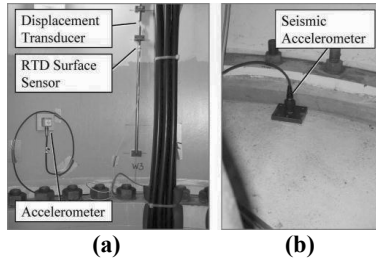
Figure 1. Illustration of sensor instrumentation: (a) Displacement transducer, RTD surface sensor, and accelerometer installed at the 42 m level inside the wind turbine tower, and (b) a seismic accelerometer at the ground level.

Data sets used in this study. The machine learning approach requires different inputs (i.e. wind field statistics) and corresponding outputs (i.e. structural response data) so that structural wind turbine behaviors are "learnt" from the existing input-output patterns. For input, long-term wind field statistics are computed from the measurements recorded by the anemometers described earlier. Specifically, the input feature vector $x = \{U_{67}, U_{67}–U_{13}, \sigma_{U67}\}$ consists of three statistical features, where U_{67} is the mean of a 20-minute wind speed time series at the 67 m height (recorded by the cup anemometer), U_{13} is the mean wind speed at the 13 m height (recorded by the ultrasonic anemometer), and σ_{U67} is the standard deviation of the 20-minute wind speed time series at the 67 m height. For the output, the vertical tower strain data, as calculated from the measurements recorded by the displacement transducer at the 42 m level (Figure 1a), are used. Quartiles of 20-minute strain time series are employed as an indirect measure for the level of fluctuation of the tower bending moment. For a classification of the quartiles, the output quartiles are discretized into 5 levels, i.e. $y = \{1, 2, 3, 4, 5\}$, on the basis of their magnitude. In total, 7,860 pairs of x (input feature vector) and y (discretized quartile levels), provided by the LCM system, serve as the basis for this study.

MACHINE LEARNING APPROACH

Wind Turbine Load Classification using Gaussian Discriminative Analysis. A wind turbine load classification function $\hat{y}(x)$ is defined by mapping the input feature vector x to the predicted wind turbine load class $\hat{y}$. The function $\hat{y}(x)$ is constructed using the Gaussian Discriminative Analysis (GDA), which is a generative learning algorithm that classifies the input features based on learned input feature distributions modeled by the Gaussian distribution. Given the trained input feature distribution $P(x|y)$ conditional on the load class y and the class prior probability $P(y)$, the posterior distribution on y given x is modeled according to the Bayes' rule as follows:

$$P(y\,|\,x) = \frac{P(x\,|\,y)P(y)}{P(x)} = \frac{P(x\,|\,y)P(y)}{\sum_y P(x\,|\,y)P(y)} \tag{1}$$

Then, for the new input feature vector x_{new}, representing the newly observed wind field statistics, the corresponding class $\hat{y}$ can be predicted according to the maximum a- posteriori detection (MAP) principle expressed as follows:

$$\hat{y} = \arg\max_y P(y \mid x_{new}) = \arg\max_y \frac{P(x_{new} \mid y)P(y)}{P(x_{new})} = \arg\max_y P(x_{new} \mid y)P(y) \quad (2)$$

In particular, GDA models the probability function $P(x|y = j)$ given a load class as a multivariate Gaussian distribution $N(\mu_j, \Sigma_j)$ (μ_j and Σ_j are the mean and the covariance matrix, respectively) and the prior probability of the load class $P(y)$ as a multinomial distribution, i.e. $P(y = j; \phi) = \phi_j$. To construct $P(x|y)$ and $P(y)$, we need to estimate the parameters μ_j, Σ_j, and ϕ_j for each load class. Using the training data sets $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$, the parameters μ, Σ, and ϕ are calculated as the ones maximizing the following log-likelihood function (Li *et al*., 2006):

$$l(\phi, \mu, \Sigma) = \log \prod_{i=1}^{m} P(x^{(i)}, y^{(i)}; \phi, \mu, \Sigma) = \log \prod_{i=1}^{m} P(x^{(i)} \mid y^{(i)}; \mu, \Sigma)P(y^{(i)}; \phi) \quad (3)$$

Joint Probability Density Estimation using Gaussian Mixture Models. The site-specific wind field characteristics are described by the wind field statistics as specified in the input feature vector x, in particular the mean and the standard deviation of wind speed data. In this study, different wind field characteristics (by day and by night) are compared using the joint Probability Density Function (PDF) $f_{X|W}(x|w)$ for input feature x conditional on the atmospheric setting w. For a certain condition w, the joint PDF for x is constructed based on the Gaussian Mixture Model (GMM), in which PDF is expressed as a linear combination of K probability density functions as follows (Figueiredo and Jain, 2002):

$$f(x) = \sum_{j=1}^{K} \phi_j f_j(x) = \sum_{j=1}^{K} \phi_j P(x \mid \mu_j, \Sigma_j) \quad (4)$$

where $P(x|\mu_j, \Sigma_j)$ is the jth Gaussian Probability Density Function (GPDF), which is given by a multivariate normal distribution $N(\mu_j, \Sigma_j)$, and ϕ_j is the weight of the jth GPDF.

To construct the Gaussian Mixture Density (GMD) function $f(x)$, the parameters μ_j, Σ_j, and ϕ_j are preliminarily estimated for each jth distribution function. To this end, the log-likelihood of the data sets $\{x^{(i)}, \dots, x^{(m)}\}$ is expressed as:

$$l(\phi, \mu, \Sigma) = \sum_{i=1}^{m} \log P(x^{(i)}; \phi, \mu, \Sigma) = \sum_{i=1}^{m} \log \sum_{z^{(i)}=1}^{k} P(x^{(i)} \mid z^{(i)}; \mu, \Sigma)P(z^{(i)}; \phi) \quad (5)$$

and the parameters μ, Σ, and ϕ can be found as the ones that maximize Eq. (5). The latent random variable $z^{(i)}$ specifies one of the k possible Gaussian distributions from

which $x^{(i)}$ is drawn. The random variable $z^{(i)}$ follows a multinomial distribution; that is, when $z^{(i)} = j$ with the probability $P(z^{(i)} = j; \phi) = \phi_j$. The feature vector x follows the jth Gaussian distribution as $P(x^{(i)}|z^{(i)}; \mu, \Sigma) \sim N(\mu_j, \Sigma_j)$. Because $z^{(i)}$, being a random variable, is not known, the estimation of the parameters based on the maximum likelihood estimation is non-trivial. The Expectation Maximization (EM) algorithm is one efficient method for estimating the parameters, given the existence of the latent random variables z.

The EM algorithm consists of two iterative steps: (1) the "E-step" evaluates the probability of $z^{(i)}$, given the current data $x^{(i)}$ and the parameters estimated in the preceding iteration; (2) the "M-step" updates the parameters ϕ, μ, Σ that maximize the log-likelihood function according to (Eq. 5) using the $z^{(i)}$ estimated in the preceding E-step (Render and Walker 1984). These two steps are repeated until the parameters converge to some acceptable values.

Expected Wind Turbine Load. The expected wind turbine load $E[\hat{y}|w]$, on condition that the atmospheric setting w can be assumed, can be expressed as:

$$E[\hat{y} \mid w] = \int_x \hat{y}(x) f_{X|W}(x \mid w) dx \approx \sum_{i=1} \hat{y}(x^{(i)}) P_{X|W}(x^{(i)} \mid w) \qquad (6)$$

Here, $\hat{y}(x)$ is the wind turbine load classification function which maps the wind input feature x onto the discretized load class $\hat{y}$, and $f_{X|W}(x|w)$ is the joint PDF for wind field input features x given a certain atmospheric setting w. Note that the wind turbine loads depend only on the wind field input features x, and that the PDF for x is subject to change depending on the atmospheric condition w. The expectation value $E[\hat{y}|w]$ thus provides the insights into how a wind turbine experiences different levels of the true load class y given an easily observable atmospheric setting w.

CASE STUDY: APPLICATION OF THE MACHINE LEARNING APPROACH USING MONITORING DATA

Construction of Joint PDFs for Wind Field Characteristics. For comparing the wind field characteristics during the daytime and the night time, 7,860 data sets are categorized into daytime and nocturnal data sets. Using four Gaussian mixtures, the joint PDF on relevant input features x is constructed for each data set using the EM algorithm. Figure 2 shows the joint PDFs of daytime and nocturnal wind field input features. The axes in the Cartesian plot denote the mean of a 20-minute wind speed time series at 67 m (U_{67}), the differences between the mean wind speed at 67 m and 13 m ($U_{67} - U_{13}$) and the standard deviation of 20-minute wind speed time series at 67 m (σ_{U67}). Therefore, each dot in the figure corresponds to the features of a 20-minute wind field. Furthermore, the probability of each input feature vector x is denoted by its color with darker color indicating a higher probability. The two PDFs as plotted in Figures 2(a) and 2(b) clearly show the different variations in the characteristics between the day and night wind field. Wind fields in the daytime have larger dispersions in each input feature space compared to nocturnal wind fields. This

is because the non-steady state boundary layer at daytime causes a more active air flow.

(a) Day   **(b) Night**
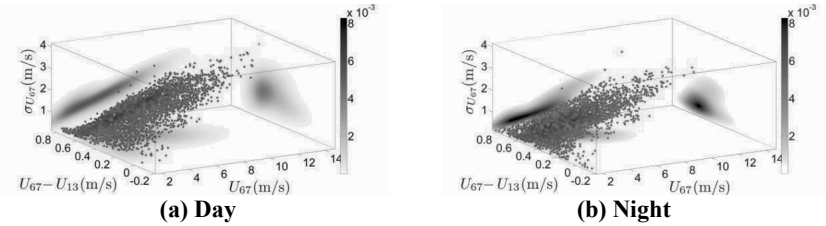
Figure 2. Joint probability density functions for the wind field input features: (a) joint PDF for the daytime, and (b) joint PDF for the night time.

Based on the modeled PDFs, the probability for any arbitrary wind input feature x can be calculated. The probabilities of wind input features are plotted in terms of three marginal PDFs as shown in Figure 3. Figures 3(a) and 3(b) show the correlation among the input features for the day time and for the night time, respectively. It can be seen that the mean and the standard deviation of the wind speed at 67 m are strongly correlated. In addition, the level of turbulence, which is depicted by the standard deviation, is lower at night than during the day, possibly because of the more stable nocturnal atmosphere.

(a) Day

(b) Night
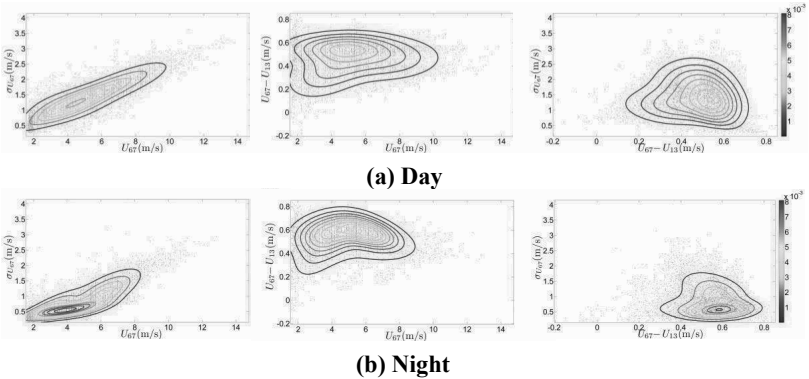
Figure 3. Marginal joint probability density functions:
(U_{67}, σ_{U67}), $(U_{67}, U_{67} - U_{13})$, and $(U_{67} - U_{13}, \sigma_{U67})$.

Classification of the Wind Turbine Response. Using the GDA approach, the wind turbine load class $\hat{y}$ is predicted. For this purpose, the quartiles of the 20-minute tower strain time series – corresponding to 3,980 sets of input features vectors – are used for training for the GDA algorithm, whereas another 3,980 sets of input features are used for testing the proposed machine learning approach. The histograms for the measured and predicted classes are shown in Figure 4. In addition, the classification errors, defined as $y - \hat{y}$, are plotted for each input-output pair. As a result, the

percentage of the exact classification ($y - \hat{y} = 0$) is about 78%. It should be emphasized that, if the error criterion is relaxed ($|y - \hat{y}| \leq 1$), the error percentage is reduced to 3.6%. Although the exact load classification is not directly needed, it is required for determining the distribution of the measured and predicted load classes, which vary due to the condition of the adjacent atmosphere and of the service of the wind turbine.
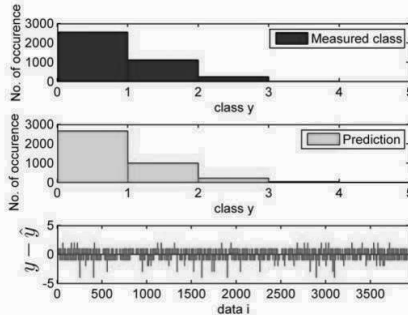
Figure 4. Wind turbine response class prediction (using the quartile range of 20-minute strain time series as wind turbine load output).

   Wind Turbine Load Prediction. The expected (time-variant) load class $E[\hat{y}|w]$ is calculated based on the input feature joint PDF and the class-mapping function. The expected classes representing the daytime and nocturnal conditions are compared in Table 1. The effectiveness of the class-mapping model is evaluated by comparing the expected class based on the measured $y^{(i)}$ and the predicted class $\hat{y}(x^{(i)})$. The comparison results show good agreement. In addition, the trend in load statistic variations can be studied by comparing the expected classes for the daytime and the nocturnal conditions. The expected load class is higher during the day than during the night, and this trend is well captured by the statistical models employed in this study.

Table 1. Comparison of the expected classes.

	Measured class $\sum_{i=1}^{N} y^{(i)} P_{X|W}(x^{(i)}|w)$	Predicted class $\sum_{i=1}^{N} \hat{y}(x^{(i)}) P_{X|W}(x^{(i)}|w)$
w = day	1.9608	2.1505 (9.7% overestimate)
w = night	1.3794	1.3021 (5.6% underestimate)

SUMMARY

   A machine learning approach towards analyzing and predicting the response data of a wind turbine structure subjected to transient wind fields is investigated. Long-term monitoring data, provided by an integrated life-cycle management framework, is used to evaluate the proposed machine learning approach. The results obtained indicate that Gaussian Discriminative Analysis (GDM) and Gaussian

Mixture Models (GMMs) can be coupled to estimate wind turbine loads in various atmospheric conditions. In the study presented in this paper, it could be observed that – according to the monitoring data used – greater wind speeds and larger standard deviations are generally observed during the day.

ACKNOWLEDGEMENTS

REFERENCES

Figueiredo, A. M. and Jain, K. A. (2002). "Unsupervised learning of finite mixture models." *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 24(3), pp. 381-396.

Hartmann, D., Smarsly, K. and Law, K. H. (2011). "Coupling Sensor-Based Structural Health Monitoring with Finite Element Model Updating for Probabilistic Lifetime Estimation of Wind Energy Converter Structures." In: *Proceedings of the 8th International Workshop on Structural Health Monitoring 2011*. Stanford, CA, USA, September 13, 2011.

Li, T., Zhu, S. and Ogihara, M. (2006). "Using discriminant analysis for multi-class classification: an experimental investigation." *Knowledge and Information Systems*, 10(4), pp. 453-472.

Render, A. R. and Walker, F. H. (1984). "Mixture Densities, Maximum Likelihood and the EM Algorithm." *SIAM Review*, 26(2), pp. 195-239.

Smarsly, K., Law, K. H. and Hartmann, D. (2011a). "Implementation of a multiagent-based paradigm for decentralized real-time structural health monitoring." In: *Proceedings of the 2011 ASCE Structures Congress*. Las Vegas, NV, USA, April 14, 2011.

Smarsly, K., Law, K. H. and Hartmann, D. (2011b). "Implementing a Multiagent-Based Self-Managing Structural Health Monitoring System on a Wind Turbine." In: *Proceedings of the 2011 NSF Engineering Research and Innovation Conference*. Atlanta, GA, USA, January 4, 2011.

Smarsly, K., Law, K. H. and Hartmann, D. (2012a). "A Multiagent-Based Collaborative Framework for a Self-Managing Structural Health Monitoring System." *ASCE Journal of Computing in Civil Engineering*, 26(1), pp. 76-89.

Smarsly, K., Law, K. H. and Hartmann, D. (2012b). "Towards Life-Cycle Management of Wind Turbines based on Structural Health Monitoring." In: *Proceedings of the First International Conference on Performance-Based Life-Cycle Structural Engineering*. Hong Kong, China, December 5, 2012.

Smarsly, K., Hartmann, D. and Law, K. H. (2012c). "Integration of Structural Health and Condition Monitoring into the Life-Cycle Management of Wind Turbines." In: *Proceedings of the Civil Structural Health Monitoring Workshop*. Berlin, Germany, November 6, 2012.

Novel sparse Bayesian learning for structural health monitoring using incomplete modal data

Yong Huang[1] and James L. Beck[2]

[1]Division of Engineering and Applied Sciences, California Institute of Technology, MC 9-94, Pasadena, CA 91125, PH (626) 395-3491; email: yonghuan@caltech.edu
[2]Division of Engineering and Applied Sciences, California Institute of Technology, MC 9-94, Pasadena, CA 91125, PH (626) 395-4132; FAX (626) 578-0124; email: jimbeck@caltech.edu

ABSTRACT

For civil structures, structural damage due to excessive loading or environmental degradation usually occurs in localized areas. A new sparse Bayesian probabilistic approach for computing the probability of localized stiffness reductions induced by damage is presented that uses noisy incomplete modal data from before and after possible damage. The methodology employs a hierarchical form of the prior for the stiffness parameters that promotes spatial sparsity in the inferred stiffness reductions. To obtain the most plausible model of sparse stiffness reductions together with its uncertainty within a specified class of models, the method employs an optimization scheme that iterates between the groups of modal parameters and hyperpameters. The approach also adopts a recent published strategy for Bayesian dynamic model updating based on modal data that has two important benefits: (1) no matching of model and experimental modes is needed, and (2) solving the eigenvalue problem of a structural model is not required. For validation, a three-dimensional braced-frame model with simulated data from the Phase II benchmark problem sponsored by the IASC-ASCE task group on structural health monitoring is analyzed using the proposed method. The results show that no threshold is required to issue a damage alarm for the proposed approach and the occurrence of false-positive and false-negative damage detection is clearly reduced in the presence of modeling error.

INTRODUCTION

Interest in developing the capability of detecting, locating and assessing structural damage accurately at an early stage in its evolution is wide spread throughout the structural engineering community. The most important objective of a damage identification algorithm is to reliably issue an alarm if damage has occurred. An alarm is generally issued if some damage features shift from their healthy state values, which usually is determined by a damage index obtained from an unknown state of the structure deviating from the healthy state beyond some threshold. Defining a proper threshold is the critical challenge to establish a timely and reliable damage alarm. Recently, some researchers have investigated how to compute a proper threshold value in a rigorous manner in order to alleviate false positive (false alarm) and false negative (missed alarm) detections (Sohn et

al., 2005; Zhou et al. 2001). However, novel methods still need to be explored for better damage alarm performance.

Another challenge for structural damage detection is that existing methods usually require measurement information at locations corresponding to every degree of freedom (DOF) of a model of the structure, whereas, in reality, sensors are typically installed at only a limited number of locations, so the spatial distribution of the structural motion is not known completely. Therefore, it is impossible to exactly describe the current state of the structure by the limited information available in practice, so we have a state of uncertainty that can be described probabilistically. Rather than consider only a point estimate for the model parameters, Bayesian inference considers all possible values of parameters and explicit treats modeling uncertainty, including quantification of parametric uncertainty, by treating the problem within a framework of plausible inference in the presence of incomplete information (Beck, 2010). The Bayesian framework has been used previously for damage detection and assessment (e.g. Vanik, 1997; Sohn and Law, 1997; Beck et al., 1998; Vanik et al. 2000; Ching and Beck, 2004; Yuen et al., 2006.).

Considering the fact that damage is a localized phenomenon that induces a spatially-sparse stiffness loss in a structure, sparse Bayesian learning (Tipping, 2001; Huang et al., 2012) is investigated to improve the accuracy and robustness of damage detection and assessment. The method employs an optimization scheme that iterates between modal parameters and hyperpameters to obtain the most plausible model of spatially-sparse stiffness reductions together with its uncertainty. For validation, the simulated data from the Phase II benchmark structure sponsored by the IASC-ASCE Task Group on Structural Health Monitoring (Bernal et al., 2002) is analyzed using the proposed method.

FORMULATION

Structural model class. For a structure of interest, we take a class of linear structural models that has N_d degrees of freedom, a known mass matrix **M** from structural drawings and an uncertain stiffness matrix **K** that is represented as a linear combination of $(N_\theta + 1)$ substructure stiffness matrices $\mathbf{K}_j$, $j = 1,2, \dots N_\theta$, as follows:

$$\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}_0 + \sum_{j=1}^{N_\theta} \theta_j \mathbf{K}_j \qquad (1)$$

where the nominal substructure stiffness matrices $\mathbf{K}_j \in \mathbb{R}^{N_d \times N_d}$ represent the contribution of the jth substructure of the structure to the overall stiffness matrix **K** and $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_{N_\theta}] \in \mathbb{R}^{N_\theta}$ are corresponding stiffness scaling parameters. The reduction of any θ_j, $j = 1,2, \dots N_\theta$, corresponds to damage in the jth substructure. Therefore, $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_d - \boldsymbol{\theta}_u$ can be considered as a sparse vector with relative few non-zero components, where $\boldsymbol{\theta}_d$ and $\boldsymbol{\theta}_u$ are stiffness scaling parameters for damaged and undamaged states. We assume that a linear dynamic model with classical normal modes is adequate for damage detection purposes because we use low-amplitude vibration to identify modal parameters before and after possible damage. Under this hypothesis, a damping matrix need not be explicitly modeled since it does not affect the model mode shapes.

Prior for system modal parameters. Assume that N_m modes of the system are to be identified that has *system* natural frequencies

$\boldsymbol{\omega} = \left[\omega_1, \ldots, \omega_{N_m}\right]^T$ and real *system* mode shapes $\boldsymbol{\phi} = \left[\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{N_m}\right] \in \mathbb{R}^{N_d \times N_m}$. We do not assume that these system modal parameters satisfy the eigenvalue problem corresponding to any structural model specified by the parameters $\boldsymbol{\theta}$ because there will always be modeling errors, so

$$\left(\mathbf{K}(\boldsymbol{\theta}) - \omega_j^2 \mathbf{M}\right)\boldsymbol{\phi_j} = \mathbf{e}_j \tag{2}$$

where the uncertain equation errors $\mathbf{e}_j \in \mathbb{R}^{N_d}$, $j = 1, \ldots, N_m$, are modeled probabilistically as independent Gaussian vectors with zero mean and covariance matrix $\sigma^2 \mathbf{I}_{N_d} = \mathrm{diag}(\sigma^2, \ldots, \sigma^2)$. A key idea first presented in Yuen et al. (2006) is that Eq. (2) is used to create the prior PDF on the system modal parameters conditional on the model parameters $\boldsymbol{\theta}$ and σ^2:

$$p(\boldsymbol{\omega}^2, \boldsymbol{\phi}|\boldsymbol{\theta}, \sigma^2) = c \cdot \exp\left\{-\frac{1}{2\sigma^2} \sum_{j=1}^{N_m} \left\|\left(\mathbf{K}(\boldsymbol{\theta}) - \omega_j^2 \mathbf{M}\right)\boldsymbol{\phi}_j\right\|^2\right\} \tag{3}$$

where c is a normalizing constant. The introduction of σ^2 in (3) allows for the explicit control of how closely the system and model modal parameters agree; as $\sigma^2 \to 0$, the system modal parameters become tightly clustered around the modal parameters corresponding to the structural model specified by $\boldsymbol{\theta}$, which are given by Eq. (2) with all $\mathbf{e}_j = \mathbf{0}$. Note also that these model modal parameters are always the most plausible values a priori of the system modal parameters.

Prior for structural model parameters. We take the prior PDF $p(\boldsymbol{\theta})$ as a Gaussian distribution for $\boldsymbol{\theta}_u$ and we choose the stiffness parameter change $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_d - \boldsymbol{\theta}_u$ to have a zero-mean Gaussian prior. We adopt the automatic relevance determination (ARD) prior (Tipping, 2001, Huang et. al, 2012) for $\Delta\boldsymbol{\theta}$ to promote sparsity, so $p(\Delta\boldsymbol{\theta}|\boldsymbol{\alpha}) = N(\Delta\boldsymbol{\theta}|0, \mathbf{A^{-1}})$ is parameterized using a diagonal covariance matrix $\mathbf{A^{-1}} = \mathrm{diag}(\alpha_1^{-1}, \ldots, \alpha_{N_\theta}^{-1})$, where each of the hyperparameters α_j^{-1} is the independent prior variance for $\Delta\theta_j$. We optimize the $\alpha_j' s$ later using the evidence framework for selecting the most probable sparse stiffness reduction.

Likelihood function. Suppose that the identified N_m modes have "measured" natural frequencies $\widehat{\boldsymbol{\omega}}^2 = \left[\widehat{\omega}_1^2, \ldots, \widehat{\omega}_{N_m}^2\right]^T$ and mode shapes $\widehat{\boldsymbol{\Psi}} = \left[\widehat{\boldsymbol{\Psi}}_1, \ldots, \widehat{\boldsymbol{\Psi}}_{N_m}\right] \in \mathbb{R}^{N_s \times N_m}$, where $\widehat{\boldsymbol{\Psi}}_i \in \mathbb{R}^{N_s}$ are components of the system mode shape of the ith mode at the measured DOF. These modal parameters are assumed to be directly estimated from dynamic data, using an appropriate modal identification procedure, such as MODE-ID (Beck and Jennings, 1980) that does not use a structural model. Based on a Gaussian model for the measurement errors, the likelihood function is:

$$p(\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}|\boldsymbol{\omega}^2, \boldsymbol{\phi}) = N\left(\left[\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}\right]|[\boldsymbol{\omega}^2, \boldsymbol{\Gamma}\boldsymbol{\phi}], \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right) \tag{4}$$

where the selection matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{N_s \times N_d}$ with "1s" and "0s" picks the observed degrees of freedom from the system mode shapes and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is a covariance matrix for the measurement error estimated from the modal identification procedure.

Posterior PDF for model parameters $[\boldsymbol{\theta}, \boldsymbol{\omega}^2, \boldsymbol{\phi}]$. The posterior PDF for model parameters $[\boldsymbol{\theta}, \boldsymbol{\omega}^2, \boldsymbol{\phi}]$ from Bayes' theorem is:

$$p(\boldsymbol{\omega}^2, \boldsymbol{\phi}, \boldsymbol{\theta}|\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}, , \sigma^2, \boldsymbol{\alpha}) \propto p(\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}|\boldsymbol{\omega}^2, \boldsymbol{\phi})p(\boldsymbol{\omega}^2, \boldsymbol{\phi}|\boldsymbol{\theta}, \sigma^2)p(\boldsymbol{\theta}|\boldsymbol{\alpha})$$
$$= p(\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}|\boldsymbol{\omega}^2, \boldsymbol{\phi})p(\boldsymbol{\theta}|\boldsymbol{\omega}^2, \boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2) \tag{5}$$
$$= p(\boldsymbol{\omega}^2, \boldsymbol{\phi}, \boldsymbol{\theta}|\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\Psi}}, \sigma^2)p(\boldsymbol{\theta}|\boldsymbol{\alpha})$$

The factor $p(\boldsymbol{\theta}|\boldsymbol{\omega}^2, \boldsymbol{\phi}, \boldsymbol{\alpha}, \sigma^2)$ in Eq. (5) is the posterior PDF for model parameter $\boldsymbol{\theta}$, which can be computed by sparse Bayesian learning with successive relaxation (Huang et al., 2012) based on the linear regression model in Eq. (2), given the value of parameters $[\boldsymbol{\omega}^2, \boldsymbol{\phi}]$:

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e} \tag{6}$$

where $\mathbf{e} \in \mathbb{R}^{N_d N_m}$ is zero-mean Gaussian with covariance matrix $\mathbf{I}_{N_m N_d}$, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{K}_1\boldsymbol{\phi}_1 & \mathbf{K}_2\boldsymbol{\phi}_1 & \cdots & \mathbf{K}_{N_\theta}\boldsymbol{\phi}_1 \\ \mathbf{K}_1\boldsymbol{\phi}_2 & \mathbf{K}_2\boldsymbol{\phi}_2 & \cdots & \mathbf{K}_{N_\theta}\boldsymbol{\phi}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{K}_1\boldsymbol{\phi}_{N_m} & \mathbf{K}_2\boldsymbol{\phi}_{N_m} & \cdots & \mathbf{K}_{N_\theta}\boldsymbol{\phi}_{N_m} \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} (\mathbf{K}_0 - \omega_1^2\,\mathbf{M})\boldsymbol{\phi}_1 \\ (\mathbf{K}_0 - \omega_2^2\,\mathbf{M})\boldsymbol{\phi}_2 \\ \vdots \\ (\mathbf{K}_0 - \omega_{N_m}^2\,\mathbf{M})\boldsymbol{\phi}_{N_m} \end{bmatrix} \tag{7}$$

The factor $p(\boldsymbol{\theta}, \boldsymbol{\omega}^2, \boldsymbol{\phi}|\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\psi}}, \sigma^2)$ in Eq. (5) is the posterior PDF over $[\boldsymbol{\theta}, \boldsymbol{\omega}^2, \boldsymbol{\phi}]$ with non-informative prior $p(\boldsymbol{\theta})$ over $\boldsymbol{\theta}$. Yuen. et al. (2006) proposed an efficient iterative scheme of linear optimizations to determine the MAP (Maximum a posteriori) estimate of $[\boldsymbol{\theta}, \boldsymbol{\omega}^2, \boldsymbol{\phi}]$ with the value σ^2 fixed at σ_{eq}^2. We denote the corresponding algorithm as Algorithm 1. We also denote the algorithm of sparse Bayesian learning with successive relaxation as Algorithm 2 here.

PROPOSED METHOD

Each time an update is performed using modal data, we initialize by running Algorithm 1 and get the most probable a posteriori system mode shapes $\widetilde{\boldsymbol{\phi}}$ with values at all DOFs and their corresponding most probable a posteriori system frequencies $\widetilde{\boldsymbol{\omega}}^2$ as well. Constructing $\mathbf{H}$ and $\mathbf{y}$ from $\widetilde{\boldsymbol{\phi}}$ and $\widetilde{\boldsymbol{\omega}}^2$, then we run Algorithm 2 to optimize the hyperparameters $[\boldsymbol{\alpha}, \sigma^2]$ to find the MAP values $[\widetilde{\boldsymbol{\alpha}}, \tilde{\sigma}^2]$ and then get the posterior PDF for $\boldsymbol{\theta}$: $\mathrm{p}(\boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}}^2, \widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{\alpha}}, \tilde{\sigma}^2) = N(\boldsymbol{\theta}|\boldsymbol{\mu_\theta}, \boldsymbol{\Sigma_\theta})$, where $\boldsymbol{\mu_\theta} = \tilde{\sigma}^{-2}\boldsymbol{\Sigma_\theta}\mathbf{H}^T\mathbf{y} + \widetilde{\boldsymbol{\theta}}_u$, $\boldsymbol{\Sigma_\theta} = (\tilde{\sigma}^{-2}\mathbf{H}^T\mathbf{H} + \mathbf{A})^{-1}$, where $\widetilde{\boldsymbol{\theta}}_u$ is the MAP value of $\boldsymbol{\theta}_u$ obtained by running Algorithm 1 for the undamaged state. We get the optimal stiffness parameter value $\overline{\boldsymbol{\theta}} = \boldsymbol{\mu_\theta}$.

To obtain the most probable values of model parameters $[\boldsymbol{\omega}^2, \boldsymbol{\phi}, \boldsymbol{\theta}]$, we need to guarantee $\mathrm{p}(\widetilde{\boldsymbol{\omega}}^2, \widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{\theta}}|\widehat{\boldsymbol{\omega}}^2, \widehat{\boldsymbol{\psi}}, \tilde{\sigma}_{eq}^2)$ has the maximum density at the location that $\widetilde{\boldsymbol{\theta}} = \overline{\boldsymbol{\theta}}$ and $\tilde{\sigma}_{eq}^2 = \tilde{\sigma}^2$, where $\overline{\boldsymbol{\theta}}$ and $\tilde{\sigma}^2$ are the most probable values updated by running Algorithm 2 given $\widetilde{\boldsymbol{\omega}}^2$ and $\widetilde{\boldsymbol{\phi}}$. We need to check this by updating $[\boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\theta}]$ again using Algorithm 1 with the nominal value $[\overline{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\omega}}^2]$ as the initial state and get new most probable $[\widetilde{\boldsymbol{\omega}}^2, \widetilde{\boldsymbol{\phi}}, \widetilde{\boldsymbol{\theta}}]$. If $\|\widetilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}}\|_2^2 / \|\overline{\boldsymbol{\theta}}\|_2^2 < 0.01$, we consider the update of $\boldsymbol{\theta}$ is converged. If not, we will run Algorithm 2 again. This iteration between Algorithm 1 and 2 is continued until convergence.

For optimization of the two parameters $\boldsymbol{\theta}$ and σ_{eq}^2, we employ a successive relaxation scheme here: first $\boldsymbol{\theta}$ is optimized until convergence with σ_{eq}^2 fixed, then we update $\tilde{\sigma}_{eq}^2 = \tilde{\sigma}^2$ with $\boldsymbol{\theta}$ fixed at its current optimal value, this procedure being repeated until convergence of the two parameters is achieved.

We produce a new Bayesian learning method for sparse stiffness reduction detection which iterates between Algorithms 1 and 2 until some specified convergence criteria are satisfied, thereby giving the most probable values and their corresponding uncertainty of the overall parameters based on the modal data. In this algorithm, the outer loop updates the prediction-error variance $\tilde{\sigma}_{eq}^2$ and is

terminated when the changes in the $\tilde{\sigma}_{eq}^2$ s are sufficiently small ; e.g. $\left\| (\tilde{\sigma}_{eq}^2)^{[l+1]} - (\tilde{\sigma}_{eq}^2)^{[l]} \right\|_2^2 / \left\| (\tilde{\sigma}_{eq}^2)^{[l]} \right\|_2^2 < 0.01$, where l is the iteration number of the outer loop. The inner loop iterates between Algorithms 1 and 2 and is terminated when $\left\| \tilde{\boldsymbol{\theta}} - \overline{\boldsymbol{\theta}} \right\|_2^2 / \left\| \overline{\boldsymbol{\theta}} \right\|_2^2 < 0.01$.

Algorithm 3. Bayesian learning for sparse stiffness reduction identification

1. **Inputs:** $\hat{\boldsymbol{\omega}}^2, \hat{\boldsymbol{\Psi}}$; **Outputs:** posterior mean and covariance of all parameters
2. Initialize model stiffness parameters $\boldsymbol{\theta}$ as the nominal values $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_u$, and the eigenvalues $\boldsymbol{\omega}^2$ as the measured values $\hat{\boldsymbol{\omega}}^2$. Set σ_{eq}^2 as some specified value for Algorithm 1 initially. Perform Algorithm 1 with non-informative prior over $\boldsymbol{\theta}$ (Bayesian updating using eigenvalue-eigenvector measurements) to get $[\tilde{\boldsymbol{\omega}}^2, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}]$.
3. **While** convergence criterion on $\boldsymbol{\theta}$ is not met (Inner loop)
4. Based on $\tilde{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\omega}}^2$, perform Algorithm 2 (Sparse Bayesian learning using successive relaxation) to get the optimal hyperparameters $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\sigma}^2$, and calculate the MAP value of $\overline{\boldsymbol{\theta}}$.
5. Initialize model stiffness scaling parameters $\boldsymbol{\theta}$ as the nominal values $\boldsymbol{\theta}_0 = \overline{\boldsymbol{\theta}}$, and the eigenvalues $\boldsymbol{\omega}^2$ as the measured values $\tilde{\boldsymbol{\omega}}^2$. Fix σ_{eq}^2 and perform Bayesian update using Algorithm 1 with non-informative prior over $\boldsymbol{\theta}$ to get $[\tilde{\boldsymbol{\omega}}^2, \tilde{\boldsymbol{\phi}}, \tilde{\boldsymbol{\theta}}]$.
6. **End while**
7. Optimize σ^2 (Tipping, 2001, Huang et.al 2012) based on the current parameter model and set $\tilde{\sigma}_{eq}^2 = \tilde{\sigma}^2$. Check if the updating of prediction error variance $\tilde{\sigma}_{eq}^2$ has converged or not. If not, fix $\tilde{\sigma}_{eq}^2$ and repeat the inner loop (steps 3 to 6); otherwise end.

RESULTS OF IASC-ASCE PHASE II SIMULATED BENCHMARK STUDY

The method is applied to a four-story, two-bay by two-bay steel braced-frame model (Figure 1) which is from the IASC-ASCE Phase II Benchmark problem. This paper uses the identified modal data $(\hat{\boldsymbol{\omega}}^2, \hat{\boldsymbol{\Psi}})$ from Ching and Beck (2004) based on the simulated benchmark data with added "measurement" noise equal to 10% RMS of the signal at the measured DOFs. A total of eight modes, four in the strong (x) direction and four in the weak (y) direction, of the structure are identified by MODE-ID (Beck and Jennings, 1980) for all cases, both for the full-sensor (measurements are available at the centre of each side at each floor) and partial-sensor (measurements at the third floor and the roof are available) scenarios, respectively. For the present example, only the results of the braces damage are presented. A detailed description of the structural properties and damage cases can be found in Bernal et al. (2002) and Ching and Beck (2004)).
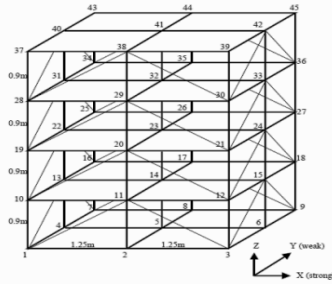
Figure 1. The diagram of the benchmark structure (Ching and Beck (2004))

The employed numerical model for damage detection is a 12-DOF shear-building with rigid floors and three DOF per floor: translations parallel to the x- and y-axes and rotation about the z-axis. The stiffness matrix K is parameterized as

$$\mathbf{K}(\boldsymbol{\theta}) = \mathbf{K}_0 + \sum_s \sum_f \theta_{sf} \bar{\mathbf{K}}_{sd} \qquad (9)$$

where $s=1,\ldots,4$ refers to the story number and $f= ' + x', ' - x', '+ y'' - y'$ indicates the face of the respective floor. Four stiffness parameters are used for each story to give a stiffness parameter vector $\boldsymbol{\theta}$ with size of 16 components.

Based on the results from the Algorithm 3, the ratio of the most probable values of the stiffness parameters θ_{sf} for different cases with respect to those for the undamaged case are tabulated in Table 1 and 2 for the two sensor scenarios, respectively. The corresponding c.o.v. for θ_{sf} are also shown, which are calculated from the ratio of the square root of the diagonal elements of the covariance matrix of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ to $\boldsymbol{\mu}_{\boldsymbol{\theta}}$.

In Tables 1 and 2, the actual damaged locations are made bold faced for comparison (Ching and Beck (2004)). We issue a damage alarm when the stiffness ratio is smaller than one, which means that the stiffness reduction ratio is larger than zero. It is seen that no false or missed damage detection alarms can be found for the full-sensor scenarios and the identified stiffness ratios are close to their actual values. For partial-sensor scenarios, there is only one false detection for $\theta_{4,-y}$ for DP2B case; however, the corresponding stiffness reduction ratio 0.99 (last column) is very close to 1. Compared with the results of the previous methods (Ching and Beck, 2004; Yuen et al., 2006), the occurrence of false damage detection in the presence of modeling error is clearly suppressed by inducing sparsity.

From the observation of corresponding c.o.v. s, the uncertainty of the estimated stiffness scaling parameter $\boldsymbol{\theta}$ at the components where no damage occurs is zero, which means these locations have no stiffness reduction with full confidence. The proposed method reduces the uncertainty for the estimated stiffness scaling parameter $\boldsymbol{\theta}$, which helps suppress the occurrence of false damage detections.

Table 1 Stiffness ratios for the full-sensor scenario in simulated Phase II.

	$\theta_{1,+x}$	$\theta_{2,+x}$	$\theta_{3,+x}$	$\theta_{4,+x}$	$\theta_{1,+y}$	$\theta_{2,+y}$	$\theta_{3,+y}$	$\theta_{4,+y}$	$\theta_{1,-x}$	$\theta_{2,-x}$	$\theta_{3,-x}$	$\theta_{4,-x}$	$\theta_{1,-y}$	$\theta_{2,-y}$	$\theta_{3,-y}$	$\theta_{4,-y}$
RB.fs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DP1B.fs	1.00	1.00	1.00	1.00	**0.88**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.89**	1.00	1.00	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	1.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.80	0.00	0.00	0.00
DP2B.fs	1.00	1.00	1.00	1.00	**0.94**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.95**	1.00	1.00	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	1.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.36	0.00	0.00	0.00
DP3B.fs	1.00	1.00	1.00	1.00	**0.88**	1.00	**0.94**	1.00	1.00	1.00	1.00	1.00	**0.89**	1.00	**0.94**	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	1.99	0.00	1.74	0.00	0.00	0.00	0.00	0.00	1.98	0.00	1.73	0.00
DP3Bu.fs	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.89**	1.00	**0.95**	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.67	0.00	1.46	0.00

Table 2 Stiffness ratios for the partial-sensor scenario in simulated Phase II.

	$\theta_{1,+x}$	$\theta_{2,+x}$	$\theta_{3,+x}$	$\theta_{4,+x}$	$\theta_{1,+y}$	$\theta_{2,+y}$	$\theta_{3,+y}$	$\theta_{4,+y}$	$\theta_{1,-x}$	$\theta_{2,-x}$	$\theta_{3,-x}$	$\theta_{4,-x}$	$\theta_{1,-y}$	$\theta_{2,-y}$	$\theta_{3,-y}$	$\theta_{4,-y}$
RB.ps	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
DP1B.ps	1.00	1.00	1.00	1.00	**0.89**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.89**	1.00	1.00	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	1.69	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.68	0.00	0.00	0.00
DP2B.ps	1.00	1.00	1.00	1.00	**0.94**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.95**	1.00	1.00	0.99
c.o.v.(%)	0.00	0.00	0.00	0.00	1.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.50	0.00	0.00	1.74
DP3B.ps	1.00	1.00	1.00	1.00	**0.89**	1.00	**0.95**	1.00	1.00	1.00	1.00	1.00	**0.88**	1.00	**0.96**	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	2.06	0.00	1.79	0.00	0.00	0.00	0.00	0.00	2.05	0.00	1.79	0.00
DP3Bu.ps	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.89**	1.00	**0.94**	1.00
c.o.v.(%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.64	0.00	1.43	0.00

CONCLUSION

A Bayesian damage detection and assessment method is proposed that uses a spatially–sparse stiffness reduction from incomplete noisy data. This new method is applied to data from the IASC-ASCE Phase II simulated benchmark. The results shows that no threshold is needed for issuing a damage alarm and the occurrence of false-positive and false-negative damage detection in the presence of modeling error is clearly suppressed.

REFERENCES

Beck, J.L. and Jennings, P.C. (1980). "Structural identification using linear models and earthquake records." *Earthquake Engineering and Structural Dynamics*, 8 (2), 145-160.

Beck, J.L. and Katafygiotis, L.S. (1998). "Updating Models and Their Uncertainties. I: Bayesian Statistical Framework." *Journal of Engineering Mechanics*, 124 (4), 455-461.

Beck, J.L. (2010). "Bayesian system identification based on probability logic." *Structural Control and Health Monitoring*, 17 (7), 825-847.

Bernal, D., Dyke, S.J., Lam, H.-F. and Beck, J.L. (2002). "Phase II of the ASCE benchmark study on SHM." In: Proceedings of the 15th Engineering Mechanics Division Conference of the American Society of Civil Engineering, 125(9), 1048-1055

Ching, J. and Beck, J.L. (2004). "New Bayesian model updating algorithm applied to a structural health monitoring benchmark." *Structural Health Monitoring,* 3 (4), 313-332.

Huang. Y, Beck J.L., Wu. S. and Li H. (2013). "Robust Bayesian compressive sensing for signals in structural health monitoring." *Computer-Aided Civil and Infrastructure Engineering* (Accepted).

Sohn H. and Law, K.H. (1997). "A Bayesian probabilistic approach for structure damage detection." *Earthquake Engineering and Structure Dynamics*, 26, (12), 1259-1281.

Sohn, H., Allen, D.W., Worden K. and Farrar, C.R. (2005). "Structural damage classification using extreme value statistics." *Journal of Dynamic Systems, Measurement, and Control,* 127(1): 125-132

Tipping, M.E. (2001). "Sparse Bayesian learning and the relevance vector machine." *Journal of Machine Learning Research,* 1, 211-244.

Vanik, M.W. (1997). "A Bayesian probabilistic approach to structural health monitoring." *Technical Report EERL 97-07,* Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA.

Vanik, M. W. and Beck, J. L. and Au, S. K. (2000). "Bayesian Probabilistic Approach to Structural Health Monitoring." *Journal of Engineering Mechanics,* 126 (7),738-745.

Yuen, K.-V. and Beck, J.L. and Katafygiotis, L.S. (2006). "Efficient model updating and health monitoring methodology using incomplete Modal Data without Mode Matching." *Structural Control and Health Monitoring*, 13 (1), 91-107

Zhou, H., Ni, Y., and Ko, J. (2011), "Eliminating temperature effect in vibration-based structural damage detection." *Journal of Engineering Mechanics*, 137(12), 785-796.

Condition Assessment of Stay Cables based on Laboratory Tests and Structural Health Monitoring

Shunlong LI[1], Yang XU[2], Hui LI[3] and Weiming YAN[4]

[1]School of Transportation Science and Engineering, Harbin Institute of Technology, P.O. Box 2537, Harbin; PH (451) 8628-3191; FAX(451)8628-2068; email: lishunlong@hit.edu.cn
[2]School of Civil Engineering, Harbin Institute of Technology, P.O. Box 2537, Harbin; PH (451) 8628-3191; FAX(451)8628-2068; email: xy_hit2008@126.com
[3]School of Civil Engineering, Harbin Institute of Technology, P.O. Box 2537, Harbin; PH (451) 8628-2013; FAX(451)8628-2013; email:lihui@hit.edu.cn
[4]Beijing Key Lab of Earthquake Engineering and Structural Retrofit, Beijing University of Technology, Beijing, 100 Pingleyuan; PH (010) 6739-2098; FAX(010) 6739-2098; email: yanwm@bjut.edu.cn

ABSTRACT

This paper presents a condition assessment approach of cables incorporating experimental deteriorated model and in-service loads collected by SHM system. The high strength steel wires were placed under simulated corrosive media based on the Salt Spray Test Standards. The geometric feathers of the corroded steel wire surface were measured by to generate the uniform and pitting corrosion depth model. The long-term deterioration process of steel wires in the cables considers simultaneously the experimented uniform and pitting corrosion due to simulated environmental attack and the fatigue propagation induced by monitored cyclic stress. By employing first order reliability method, the reliability of the cables under the monitored responses is further estimated in terms of the serviceability specified in the design specification. The life-cycle condition assessment of stay cables provide guidance to the future decision making related to maintenance and replacement.

INTRODUCTION

For long-span cable stayed bridges, the cables are always a critical and vulnerable type of structural components in normal operation conditions. During the operation, the corrosive erosion, material aging and traffic loads will inevitably lead to the damage accumulation and resistance deterioration of stay cables, which could decrease its natural disasters resistance and service ability, even lead to trigger catastrophic accidents at extreme cases.

During the past decades, great efforts have been taken to the condition assessment of stay cables. For example, Takena & Miki (1992) presented fatigue tests

of a parallel-wire large diameter cable system and conducted a detailed review of other fatigue results. Matteo & Deodatis (1994) and Betti & West (2005) presented the methodology to estimate the current safety factor of the main suspension cable based on the laboratory testing of wire samples extracted from real cables for the tensile strength and elongation. Very few studies have been reported on the conditions assessment of stay-cables incorporating laboratory testing and Structural Health Monitoring (SHM) technique, which is essential information to optimize their inspection and replacement in real bridge maintenance practice.

This paper presents an investigation of the long-term condition assessment of cables based on the installed SHM system. First, the deteriorated models, including distributions of uniform corrosion, pitting corrosion depth with time were measured by laboratory environmental simulation. Second, the stochastic deterioration process of steel wires under actual situations is simulated using experimental results and monitored stresses, in which a coupled corrosion fatigue process of steel wires involving uniform corrosion, pitting corrosion and cyclic fatigue is taken into account. Third, the time-variant conditions of the cable due to corrosion fatigue is presented, including the ratio of broken wires, the distribution of crack depth and resistance of the cable. Subsequently, the reliability indices of the cable are evaluated in terms of serviceability criteria.

EXPERIMETAL DETERIORATED MODELS OF STEEL WIRES

Corrosive Conditions. The salt spray test chamber YWX/Q-016, shown in Figure 1, produced by Jiangsu Emerson Test Equipment Technology Co., LTD. was used for the corrosion testing of steel wires. The corrosion test is designed according to Salt Spray Test Standard ISO9227:1990 and the copper-accelerated acetic acid-salt spray (fog) test (CASS) is employed in this study. The following is a summary of the conditions employed:

Figure 1. Salt Spray Test Chamber YWX/Q-016.

Fog solution: 5±0.5wt% NaCl, 0.26±0.02 g/L $CuCl_2.2H_2O$, pH = 3.0~3.1, adjusted with acetic acid; Temperature: 35±1°C; Erosion periods: 24h, 48h, 72h, 96h, 144h, 168h, 240h, 480h, and 720h according to Salt Spray Test Standards ISO 9227:1990.

Sample Preparation. The raw materials of the test are 61×7mm stay-cables produced by Liuzhou OVM Machinery Co., LTD. The composition of steel wires was 0.75~0.85% carbon, 0.6~0.9% manganese, 0.12~0.32% silicon, less than 0.025% sulfur and 0.025% phosphorus.

There are 32 specimens with approximately 250mm in length for each erosion period. The severity of corrosion was assessed by observing the general corrosion of a set of 250mm. Samples exposed to the erosion periods given above were periodically removed and tested for mass loss by oxidation according to testing procedures in Standards ISO 9227:1990.

Uniform Corrosion. The corrosion products on the specimen's surface were removed by hydrochloric acid solution (volume ratio 1:1 = water:HCl mixed with 3.5g/L Hexamine, HCl density ρ_{HCl}=1.17g/ml). Mass loss of specimens was measured and this mass was subtracted from the original mass as well as the mass loss of referenced specimen to determine mass loss. The general corrosion depth of the steel wires can be calculated by the following equation:

$$v_k = \frac{W_{0k} - W_k - \dfrac{1}{n}\sum_{i=1}^{n}(W_{0i} - W_i)}{\rho\pi D L_w}$$

where v_k represents the uniform corrosion of the kth specimen; W_{0k} and W_k indicate the original mass and that removing corrosion products after experiment; ρ is the density of zinc or steel, according to the corrosion status; D, L_w refer to the diameter and length of specimen respectively; W_{0i} and W_i are the mass before and after the ith referenced specimen, used for determining mass loss caused by hydrochloric acid solution. The referenced specimen would be immerged into the hydrochloric acid solution together with experimental specimen during the corrosion products remove procedure. In the above equation, the diameters of specimens were considered to remain steady during the corrosion process, for the diameter changes were relatively small compared to the diameters itself. The statistical properties, probability density of uniform corrosion with time is shown in Figure 2.
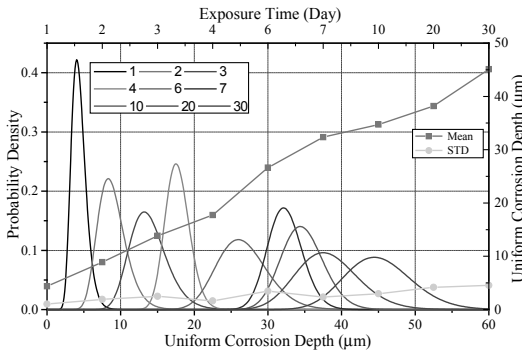
Figure 2. Uniform corrosion results.

The uniform corrosion of zinc coating could be simulated by the following equation

$$\upsilon(t) = \begin{cases} \psi_1 t^{\gamma_1} & t \le t_0 \\ \psi_1 t_0^{\gamma_1} + \psi_2(t-t_0)^{\gamma_2} & t > t_0 \end{cases}$$

where $\upsilon(t)$ is the corrosion depth at time t (in days); t_0 represents the time when zinc layer is consumed and steel starts to corrode; $\psi_1 = 4.277\mu\text{m/day}, \gamma_1 = 1.04$ and $\psi_2 = 0.398\mu\text{m/day}, \gamma_2 = 1.10$ are model parameters for zinc coating and steel respectively generated from environmental results in this study. ψ_1 and ψ_2 are proved to follow lognormal distribution whose expected value and variation coefficient are equal to ($4.277\mu\text{m/day}, 0.165$) and ($0.398\mu\text{m/day}, 1.069$) respectively. Figure 3 illustrates the regression model for the uniform corrosion together with the 95% bound, showing that the above regression model could predict the uniform corrosion depth in a certain confidential interval.

The approximate zinc coating corrosion rate in actual environmental conditions could be predicted by using the on line software http://www.galvinfo.com:8080/zclp/. For the given location of the cable stayed bridge, where the average temperature (15.6°C in this study), relative humidity (73%), precipitation (1047mm/year), deposition rate of NaCl (53mg/m^2.day, coastal environment) and concentration of SO^2 in air (SO^2 and NO^2, $88\mu\text{g/m}^3$) would be employed, the annual corrosion rate of zinc coating (6.1μm/year) could be calculated. Based on the corrosion rate and the regression model, the current cycle day gives 0.7 year/day for the current corrosive conditions.
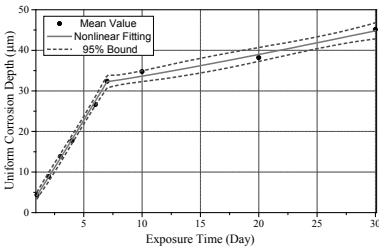
Figure 3. Uniform corrosion regression model of zinc coating.
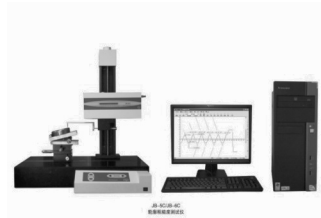
Figure 4. Roughness Measuring Instrument.

Surface Measuring. In order to describe the corrosion state of specimens and investigate the surface characteristics of steel wires, JB-5C surface roughness instruments (as shown in Figure 4), manufactured by Shanghai Tai Ming Optical Instrument Co., LTD, was employed to measure corroded surfaces of steel wires after salt spray test. Each circle of specimen was measured every 12°, and totally 30 surface lines could be collected to describe the whole surface as shown in Figure 5.

Pitting Corrosion Model. According to St. Venant's Principle the minimum element length can be no less than two steel wire diameters, so that a pit reduced capacity for the entire element (on the basis of brittle fracture being the mode of

failure). In this study, the element length of steel wires was selected as no less than 3 steel wire diameters ($L_\kappa = \kappa \times 7\text{mm}, \kappa \geq 3$).
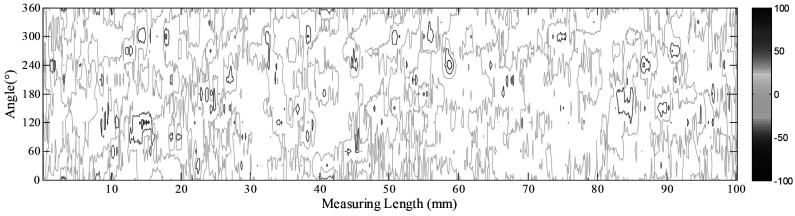
Figure 5. Typical contour outlines of specimen at 240h, 480h and 720h.

The whole measured surfaces of specimens were equally divided into L / L_κ lengths. Maximum pit-depths $a(t)$ were measured for each selected element length of the steel wires at 480h and 720h. The ratio of maximum penetration of pitting $a(t)$ to uniform corrosion depth $v(t)$, i.e. $\chi = a(t)/v(t)$, follows the Extreme Value Type I distribution, and its cumulative distribution function can be expressed as

$$F_\kappa(\chi) = \exp\left\{-\exp\left[-\frac{(\chi - \beta_\kappa)}{\alpha_\kappa}\right]\right\}$$

where α_κ and β_κ are the distribution parameters for L_κ element length. Figure 6 gives the probability density and cumulative density of measured maximum pitting factor at different element elements. It could be concluded from Figure 6 that the parameter α_κ of Extreme value distribution remains nearly the same.
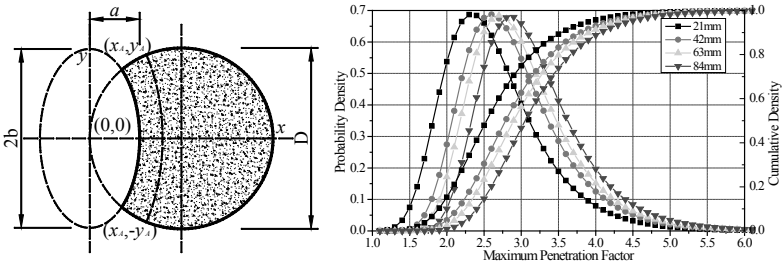
Figure 6. Pitting confiagation and distribution of maximum pitting factor.

Suppose that the distribution parameters for wires with any given length can be expressed as

$$F_\kappa(\chi) = \left[F_3(\chi)\right]^\vartheta$$

then

$$\beta_\kappa = \beta_3 + \frac{\ln(\vartheta)}{\alpha_3} = \beta_3 + \frac{\xi \ln(A_k / A_3)}{\alpha_3}, \alpha_\kappa = \alpha_3$$

where A_k represents the surface area of a given high strength wire; A_3 is the surface area of a wire with 21 mm length and 7 mm diameter, $\alpha_3 = 0.531$, $\beta_3 = 1.336$. Figure 7 gives the relationship between $\ln(A_k / A_3)$ and model parameter β_κ. Thus the parameter $\xi = 0.197$ would be generated by the measured maximum pitting factor and the distribution of any given element length could be calculated by the above regression model.
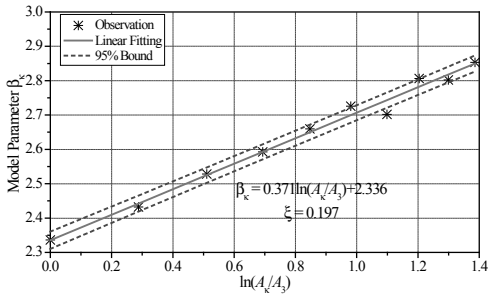
Figure 7. Regression model of $\ln(A_k / A_3)$ and β_κ.

CORROSION FATIGUE AND RELIABILITY ASSESSMENT OF CABLES

The corrosion fatigue crack growth of corrode wires was estimate at the central (deepest) point of the crack front. It is known that for the environmental-assisted fatigue situations, the cracks grow only when the stress intensity factor range ΔK is larger than corrosion fatigue threshold (in this study $2.8 \text{MPa} \cdot \sqrt{m}$). According to the experimental results of Martin & Sanchez (1990), the Paris-Erdogan law can be used for the corrosion fatigue crack growth evaluation.

$$\frac{da}{dN} = C\Delta K^m$$
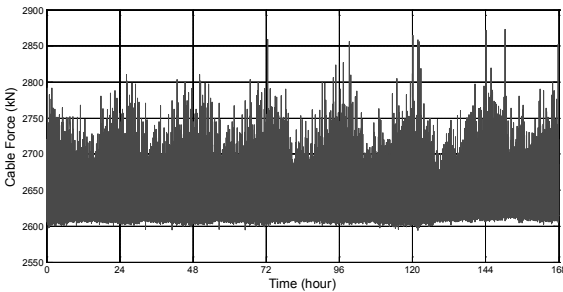
Figure 8. Time history of axial force responses in one week (Mar. 1-7, 2012).

where da / dN indicates the crack growth rate; m represents the Paris Exponent; $\Delta K = K_{max} - K_{min}$ demonstrates the stress intensity factor range, C is crack growth rate. There is a strong correlation between C and m, and their values are highly depending

on the environmental aggressiveness. Their results indicated that the coefficient m in the Paris law is around 3 and can be treated as a constant, whereas the crack growth rate of high strength wire, C, has much more randomness and follows the lognormal distribution (its mean value and variation coefficient are equal to 5.50×10^{-12} and 0.1 respectively.

The monitored cable forces results indicate that the cable A07 has the maximum stress range and would be identified as the most critical cable presented hereinafter. Figure 2 illustrates one-week cable force time history collected by SHM system.

Table 1. Basic parameters of selected cable A07.

Model	Length(m)	Area(cm^2)	Mass(kg/m)
151Φ7	150.24	58.112	45.618

Figures 3 (a) show the statistical results under repeated traffic load actions for steel wires of 7.00 mm. Figure 3 shows the distribution of corrosion fatigue crack depth after 63 year operation time, and Kolmogorov Smirnov (K-S) Hypothesis Testing at significance level of 5% indicates that the distribution does not reject the Lognormal Distribution.
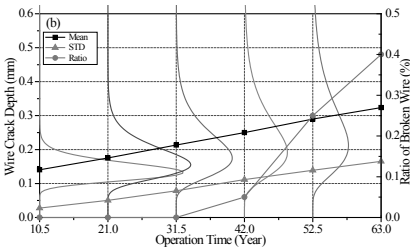
Figure 9. Evolution of wire crack depth and broken wire ratio.
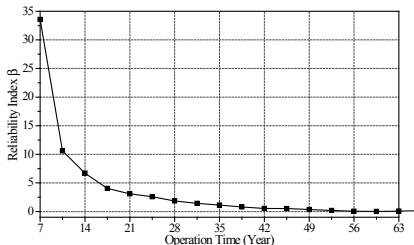
Figure 10. Long-term serviceability reliability analysis.

According to China's Technical Code of Maintenance for City Bridge (J281-2003), the cable in cable-supported bridges should be replaced when the loss of cable's cross sectional area caused by corrosion exceeds 10%. This criterion is taken as the serviceability limit state of the cable examined in this paper. Based on the time-variant distributions of cross sectional area loss of the cable due to corrosion fatigue, the long-term reliability corresponding to the serviceability is assessed. Figure 4 illustrates the evolution of the reliability index in terms of the cable's serviceability during the bridge design life. The reliability index drops considerably in the early service period, while the deterioration speed would become slow in the subsequent service period. Seen from Figure 4, the reliability index of the serviceability of the cable is less than 2 after 30-year service. As a result, the maintenance and replacement of the cables will be likely governed by the serviceability criterion if applied.

CONCLUSION

The life-cycle reliability of the cable in terms of the serviceability (the loss of cross sectional area) is computed and discussed based on the experimental corrosion model and SHM monitored stress of the cable. The serviceability limit state is defined with regard to the ratio of broken wires. The serviceability reliability analysis implies that the periodic maintenance and replacement is necessary in the whole service life. The application of proposed method to the in situ cable proved the correctness and effectiveness.

In practice, the stochastic structural response under actual environment conditions is often difficult to be characterized and modeled without in situ measurement. The presented approach to incorporate real-time SHM data as well as experimental corrosion model into condition assessment would lead to a better description of the bridge in the service life.

ACKNOWLEDGEMENTS

REFERENCES

Betti, R., West, A.C. and Vermaas, G. (2005). "Corrosion and Embrittlement in High-Strength Wires of Suspension Bridge Cables." *Journal of Bridge Engineering*, 10(2), 151-162.

Mayrbaurl, R.M., and Camo, S. (2001). "Cracking and Fracture of Suspension Bridge Wire." *Journal of Bridge Engineering*, 6(6), 645-650.

Matteo, J., and Deodatis, G. (1994). "Safety analysis of suspension bridge cables: Williamsburg Bridge." *Journal of Structural Engineering*, 120(11), 3197-3211.

Rusk, D.T., and Hoppe, W. (2009). "Fatigue life prediction of corrosion damaged high strength steel using an equivalent stress riser (ESR) model Part I: Test development and results." *International Journal of Fatigue*, 31, 1454-1463.

Sih, G.C., and Tang, X.S. (2008). "Fatigue crack growth behavior of cables and steel wires for the cable-stayed portion of Runyang Bridge: Disproportionate loosening and/or tightening of cables." *Theoretical and Applied Fracture Mechanics*, 49, 1–25.

Takena, K., Miki, C. (1992). "Fatigue resistance of large-diameter cable for cable stayed bridges." *Journal of Structural Engineering*, 118(3), 701-715.

Xu, J., Chen, W.Z., and Liu, X. (2008). "Deterioration mechanism of cables and mechanics model of wires." *Journal of Tongji University (Natural Science)*, 36(7), 911-915 (In Chinese).

Systems Modeling Approach for Sustainable Infrastructure

Islam EL-ADAWAY[1]

[1]Assistant Professor: Civil and Environmental Engineering, Mississippi State University, 501 Hardy Road, 235C Walker Engineering Building, P.O Box 9546, Mississippi State, MS 39762. E-mail: eladaway@cee.msstate.edu.

ABSTRACT

Management of our infrastructure systems is not only associated with engineering and economic constraints; it is intricately tied to public policy and institutional control considerations. As such, using simulation models that incorporate aspects of population movements, infrastructure assessment, living conditions, and health data with spatial data modeling, and collectively relate them to the different sustainability indicators could be an ideal tool for providing better information for the sustainability of our infrastructure systems. This paper presents a systems modeling approach to assess the social, environmental, and economic changes within community host systems as a result of civil infrastructure projects, and provide better information to associated stakeholders. In previous research, the author developed three novel benchmarks (i.e. nature, work, and flow) to holistically analyze our infrastructure systems from a sustainability perspective. The "work benchmark" defines the behavioral relationships between the construction processes and the associated stakeholders; the "nature benchmark" studies the interactions between the construction processes and the surrounding ecosystems, and the "flow benchmark" will analyze the overall system change. In this paper, the author shows how the "flow benchmark" will be used to demonstrate technological integration between macro-level system dynamics modeling, micro-level agent-based simulation, and multi-objective optimization to measure the overall system change. Integrating these techniques in a single framework will combine the strengths of the three methodologies and mitigate their deficiencies on stand-alone basis. This aggregation will evaluate societal vulnerability through analyzing which sustainability indicators have the most significant correlation to the affected infrastructure systems. This will create a knowledge base of simulated dataset to understand the micro-level behaviors of community stakeholders in relation to the societal macro-level sustainability outcomes. The envisaged system will analyze qualitatively and quantitatively the various sustainability indicators to help attain Pareto optimal resource balance in our communities. When successful, this innovative systems framework will provide life-cycle analysis method for estimating the probability distributions of financial cost, loss of service and societal costs, and environmental impacts. This will answer important management questions such as "what should we do?", "what if we do?', and "what are the expected societal consequences?", which will promote informed management and decision making.

INTRODUCTION

While sustainability indicators used to assess infrastructure systems usually include measures of the social, environmental/health, ecological, and economic impacts; sustainability metrics are often so general that the important system details are lost because they focus on snapshots of sustainability rather than monitoring changes in sustainability over time (Guikema 2011). Coupled with these limitations is the disconnection between communities and their host systems (Hendrickson et al. 2006). A myriad of global economic factors have forced questionable land use practices to accommodate both capital mobility and the persistent risk of wholesale capital flight (LeRoy and Mcilvaine 2010). Additionally, recent urban governance trends - that promote regionalism or liberal expansionism - have even exacerbated that many urban core issues cannot adequately be addressed by regional level resources and are rather contextualized within broader geographic, social, economic, and political frameworks (Imbroscio 2010). To this end, while many valuable simulation models have been developed to help obtain higher quality data for subsequent analysis and decision making for infrastructure management, these models are set-off actual social data and lack human-environment system components (Blaikie et al. 1994; Cannon 1994; McEntire 2001; Kunreuther 2002; Kasperson et al. 2003; Turner et al. 2003). Thus, effective and efficient management of our infrastructure systems is not only associated with engineering and economic constraints; it is intricately tied to public policy and institutional control considerations. As such, using simulation models that incorporate aspects of population movements, infrastructure assessment, living conditions, and health data with spatial data modeling, and collectively relate them to the different sustainability indicators should be ideal tool for providing better information on societal vulnerability and impact (Burnham et al. 2012). This system modeling approach will promote informed management and decision making.

GOAL AND OBJECTIVES

This paper will present the theoretical foundation and demonstrate technologies to create a knowledge base of simulated dataset to comprehend the micro-level behaviors of community recovery and resilience efforts in relation to the societal macro-level sustainability outcomes. The author hypothesizes that is that utilizing a systems-based approach, which incorporates public polices and institutional controls, can better study the interdependencies, interactions, and correlations between the different sustainability indicators of our communities. Ultimately, this research will bridge between engineering and social sciences to better understand the ability to act, reason to act, and knowledge of what to do within the context of civil infrastructure management.

BACKGROUND INFORMATION

For any civil infrastructure project, the interconnected and interdependent variables associated with the construction ecosystem are affected by rules and regulations, and are shaped by the ever-changing and developing nature of the actors, settings, and resources. According to Allen et al. (1982); Begon et al. (1990), and

Bromley (1991), infrastructure ecology is the theory and practice of evaluating the infrastructure systems within our built environment in a manner similar to the natural and social sciences where an integrated approach is used to investigate how associated stakeholders act and interact with one another as a part of a larger community host system. Also, according to Haapio and Viitaniemi (2008), systems theory is determining the effectiveness and efficiency of a system (whether the system is a product, process, or human activity) using the careful and complete assessment and evaluation of a range of factors from engineering, social science, and humanities, and their influences on each other. To this end, and based on the concepts of infrastructure ecology and systems theory, the author developed three novel benchmarks to holistically analyze the environmental, social, and affected infrastructure systems. The "work benchmark" defines the behavioral relationships between the construction processes and the associated stakeholders in the built environment to study the social and economic sustainability. The "nature benchmark" defines the interactions between the construction processes and the surrounding ecosystems to study the environmental and economic sustainability. Finally, the "flow benchmark" analyzes the overall system change. While the more interested reader can check the scientific and logical basis of the "nature" and "work" forms under El-adaway and Knapp (2012), this paper will define and detail the modeling and simulation works associated with the "flow benchmark".

THE FLOW BENCHMARK

The "flow benchmark" will predict the overall system change through a dynamic over-time simulation of the civil infrastructure resource dynamics. In other words, this benchmark will analyze interdependencies, interactions, and correlations between social, environmental, and economic sustainability indicators within community host systems. As such, the "flow benchmark" lends itself to modeling and simulation techniques that should qualitatively and quantitatively analyze the various sustainability indicators, and should help attain Pareto optimal resource balance within community host systems

Literature Review

System dynamics is a continuous simulation methodology that uses concepts from engineering feedback control theory to model and analyze socio-economic systems (Sterman 2000). The mathematical description is realized with the help of ordinary differential equations. The structure consists of multiple interacting feedback loops that depict the policies and continuous processes underlying discrete events. Simulation is governed entirely by the passage of time and is referred to as time-step simulation. At the end of each step, some system variables (i.e. level, rate and auxiliary variables) are brought up to date for representing the resulting effects of previous simulation steps (Fisher at al. 2000).

The focus on feedback loops and time delays makes system dynamics a valuable tool for investigating infrastructure resource dynamics. One important advantage of system dynamics is the possibility to deduce the occurrence of a specific behavioral mode because the structure that leads to systems' behavior is explicitly transparent (Fisher at al. 2000). The drawback though of using system dynamics is

that this structure has to be determined before starting the simulation (Sterman 2000). For instance, if a flexible system is to be modeled, every possible participant has to be included into the model and linked to the other system participants in advance. Therefore, integration of system dynamics with agent-based modeling can significantly reduce the a priori complexity of the model.

Agent-based modeling, a methodology stemming from the field of social science, simulates systems comprised of multiple idiosyncratic agents to capture the fine-grained level social interaction. The basic building block of a system is the individual agent. An agent is an autonomous program acting independently but on behalf of another actor within a system (Vlassis 2003). Agents are governed by some set of rules or behavior patterns that effect a change on the agent itself and its outside environment including other agents (North and Macal 2007). Agents typically do not work in isolation, but combine to form systems that interconnect separately developed agents and thus enable the ensemble to function beyond the abilities of any singular agent in the system (Weiss 1999). Accordingly, if a problem domain is particularly complex, large, or dynamic; it can be adequately addressed through developing agents that specialize in solving certain domain problems (Vidal 2007). To study the entire system, various modeling issues need to be properly managed including: (1) control, to know what hierarchic relationships exist between the modeled agents and how are they synchronized; (2) communication, to know what kind of messages do they send each other, and what syntax do these messages obey; and (3) learning, to know what decision-making mechanisms are available to the agents and what is the link between their perceptions, representations, and actions (Ren et al. 2003). Through control, coordination, and learning, the issues of individualism versus holism is addressed where the behavior of an agent will be determined by its cognitive structure and the dynamics of the system will arise from the interactions of agents. Thus, agent-based modeling can examine infrastructure resource dynamics to analyze the interactions of the associated stakeholders with each other using their specific internal decision structures that are guided by engineering constraints, public policies, and institutional controls.

Integrating system dynamics and agent-based modeling in a single framework will combine the strengths of the two methodologies and mitigate their deficiencies on stand-alone basis to closely understand the casual interdependencies that allow improved community planning. This integration has been successfully used previously in wide operation management applications (Nilsson and Darley 2006).

Development Principles

The proposed framework will be developed through molding three levels of aggregation. The "macro-level" will utilize system dynamics to model the stakeholders' continuous flow of the different community resources within the associated system modules (i.e. population, agriculture, industry, tourism, water-resources, pollution-control, water-quality, forest, coasts, fishery, investment … etc). At any specific point during a simulation run, the preferred structure of the resource dynamics is determined by the interactions between the associated stakeholders. Thus, the "micro-level" will detail a network of agents representing the various stakeholders associated with the decision making and management process of infrastructure systems (i.e. different community residents, government, investors,

insurers, operators… etc). Every link between two agents can be interpreted as a potential community relationship. Each of these agents will be governed by a cognitive structure that will make agents learn to choose certain decisions over others. These decisions represent the internal structure of each agent and are based on engineering constraints, public policies, and institutional controls. As a result of agents' cognition, some relationships will be active and others will be idle during the simulation process. Agents' rules will be modeled using the feedback and learning algorithms to enable the agent to adapt to its environment over time using norms, values, and beliefs.

Finally, the author will also use "multi-objective optimization" to allow decision-makers seek the Pareto optimal sustainable resource balance among alternative strategies and to use ranked prioritization towards welfare uniformity in society.

Macro-Level System Dynamics Modeling

The author will develop a macro-level system dynamics model to study the system's flow of resources within community systems. To this end, examination of the associated system modules is essential for effective and efficient deductive analysis of the system feedbacks. Thus, the macro-level modeling will accommodate various subsystems such as population, agriculture, industry, tourism, water-resources, pollution-control, water-quality, forest, coasts, fishery, investment … etc. The objective of integrating all these system modules together in one model is to simulate the implications of different strategies on the system through its subsystems. Thus, interactive relationships among the subsystems can be clarified and the embedded links among a number of system components can be identified. This should help integrating engineering constraints, public policies, and institutional controls to promote informed management and decision making to governments and humanitarian organizations.

Micro-Level Agent-Based Modeling

The author will develop a micro-level agent-based model to study the broad sustainability interdependencies of infrastructure systems within community systems. This model will incorporate two analytic methods. The first method will emphasize the social networks of the system where people affect and drive resource usage after (Begon et al. 1990). Thus, the author will simulate exchanges of information and services between the agents. Accordingly, the evolution of the system can be very sensitive to the structure and dynamics of social networks. As shown, this method is closely tied with the aforementioned macro-level system dynamics modeling. The second method will emphasize how an agent develops and then acts on its own representation of a resource to transform this resource to other agents (Bromley 1991). Thus, the author will simulate the problem of managing common renewable resources by examining the different representations of coordination actions that affect the resources in question using the various sustainability indicators. These two methods will: (1) provide a framework in which applied models take into account both the agents' interactions and systems' exchanges through the affected communities, and (2) create an extremely useful tool for decision-makers wishing to analyze policy changes or planning strategies likely to mitigate the negative effects of civil infrastructure development.

Multi-Objective Optimization

Experience has shown that simulation models are inadequate to optimize the performance of a complex system due to the large amount of development and running time required (Eksioglu 2010). Simulation-optimization models, on the other hand, iteratively identify sequence of system configurations so to reach to an optimal or near optimal solution. To this end, the author will develop multi-objective optimization algorithms (i.e. tabu search, genetic algorithms, or scatter search) to allow decision-makers seek the Pareto optimal resource balance in affected community systems to attain welfare uniformity in society.

FUTURE WORK

The author is currently working on another research that is utilizing the "nature" and "work" benchmarks to determine spatial interdependencies, interactions, and measurements of the different sustainability indicators to verify which factors and characteristics are interrelated and the degree of and/or types of risks faced by our societies. Once this process is completed, it will be integrated with the aforementioned "flow benchmark". This holistic framework will set parameters and rates of change through: (1) measuring the five sustainability indicators (i.e. environmental, social/cultural, economic, technical, and individual) for given community systems which is the subject of another ongoing research, and (2) studying the entire system and subsystems performance over a given time period using vector forms representing the probability density function of the different flow resource models.

CONCLUSION

The "flow benchmark" demonstrates technological integration between macro-level system dynamics modeling, micro-level agent-based simulation, and multi-objective optimization to measure the overall system change. This aggregation will evaluate societal vulnerability through analyzing which sustainability indicators have the most significant correlation to the affected infrastructure systems. This will improve community recovery and resilience efforts through incorporating regional public policies and institutional controls together with the traditional analysis of engineering and economic factors. This approach will fill in the gap within the available sustainability tools and rating systems that are set-off from actual social data and lack human-environment system components. It will also create a knowledge base of simulated dataset to understand the micro-level behaviors of community recovery and resilience efforts in relation to the societal macro-level sustainability outcomes. When successful, this innovative systems framework will allow: (1) creating true physical and engineering models for infrastructure systems; (2) implementing life-cycle analysis to estimate the resource flow models across all effected community systems; and (3) developing optimal resource sharing strategies that are in line with sustainability indicators' attributes. This will answer important

management questions such as "what should we do?", "what if we do?', and "what are the expected societal consequences?".

REFERENCES

Anumba, C. J., O. O. Ugwa, and Z. Ren (2005), *Agents and Multi-agent Systems in Construction*, Taylor & Francis, New York.

Begon M., Harper J.L., and C.R. Townsend (1990), *Ecology. Individuals, Populations and Communities*. Blackwell, Cambridge

Bromley D.W. (1991), *Environment and Economy, Property Rights and Public Policy*, Blackwell, Cambridge.

Coelho. Roberta de Souza, Uirá Kulesza, Arndt von Staa, Carlos José Pereira de Lucena (2006), "Unit Testing in Multi-agent Systems using Mock Agents and Aspects", *Departamento de Informática*, Pontifícia Universidade Católica Do Rio De Janeiro, Rio De Janeiro, Brazil.

DeAngelis, D. L. and L. J. e. Gross (1992), *Individual-Based Models and Approaches in Ecology*, Chapman and Hall.

Eksioglu, S. (2011), Models for Supply Chain Design and Logistics Management of Biofuels, Mississippi State University, United States.

El-adaway, Islam and Jerome Knapp (2012), "Towards a More Sustainable Rating for the Built Environment: A Proposal for Comprehensive Construction Assessment Framework", *2012 Construction Research Congress,* American Society of Civil Engineers, Indiana, United States, pp. 1981-1990.

Elmualim, A.A., Czwakiel, A., Valle, C.R., Ludlow, G. and Shah, S. (2009) "The Practice of Sustainable Facilities Management: Design Sentiments and The Knowledge Chasm*", Architectural Engineering and Design Management*, 5 (Special issue 1-2: Design Management for Sustainability). pp. 91-102.

Fenves, S., U. Flemming, C. Hendrickson, M.L. Maher, R. Quadrel, M. Terk and R. Woodbury (1996), *Concurrent Computer-Integrated Building Design*, Prentice-Hall, New Jersey, USA.

Ferber, J. (1999), *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence.* Reading, MA, Addison-Wesley.

Foley, B.A., Daniell, T.M., Warner, R.F., (2003), "What is Sustainability and Can it be Measured?*", Australian Journal of Multidisciplinary Engineering*, Vol 1, No. 1, pp.1-8.

Guikema, S. (2011), Integrated Modeling of Sustainability and Reliability for Interdependent Infrastructure Systems, John Hopkins University.

Haapio, A. and P. Viitaniemi (2008), "A Critical Review of Building Environmental Assessment Tools", *Environmental Impact Assessment Review*, Vol. 28, No. 7, Pages 469-482.

Hendrickson, C. T., Lave, L. B., and Matthews, H. S. (2006), *Environmental life cycle assessment of goods and services: An input-output approach*, RFF Press, Washington, DC.

Hogeweg, P. and B. Hesper (1990), Individual-Oriented Modeling in Ecology. *Mathl. Comput. Modeling*, 13: 83-90.

Imbroscio, D. (2010), *Urban America Reconsidered: Alternatives for Governance and Policy*, Cornell University Press, New York.

Kunreuther, H. (2002). Risk Analysis and Risk Management in an Uncertain World. *Risk Analysis, 22*(4), 655-664.

Nilsson, F. and V. Darley (2006), "On Complex Adaptive Systems and Agent-based Modeling for Improving Decision-Making in Manufacturing and Logistics Settings", *International Journal of Operations & Production Management,* Vol. 26 No. 12, pp. 1351-1373

Phelan, S. E. (1999), "A Note on the Correspondence between Complexity and Systems Theory", Systemic Practice and Action Research, Vol. 12, No. 3, pp. 237–246.

Ren, Z, C. J. Anumba, and O. O. Ugwa (2003), "Negotiation In multi-agent Systems for Construction Claim Negotiation", *International Journal of Applied Artificial Intelligence*, Volume 16, No 5, pp 359-394.

Schieritz, N. and A. Größler (2003), "Emergent Structures in Supply Chains - A Study Integrating Agent-Based and System Dynamics Modeling", *Proceedings of the 36th Hawaii International Conference on System Sciences*.

Scholl, H. J. (2001-a), "Agent-based and System Dynamics Modeling: A Call for Cross Study and Joint Research", *Proceedings of the 34th Hawaiian International Conference on Systems Science*, Wailea.

Scholl, H. J. (2001-b), "Looking Across the Fence: Comparing Findings From SD Modeling Efforts With those of Other Modeling Techniques", *Proceedings of the 19th International Conference of the System Dynamics Societ*y, Atlanta.

Sterman, J. (2000), *Business Dynamics: Systems Thinking and Modeling for a Complex World,* McGraw-Hill.

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L. et al. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences, (100),* 14, 8074-8079.

Vidal, J. M. (2007), *Fundamentals of Multi-Agent Systems*, Department of Computer Science and Engineering, University of South Carolina, Columbia, USA.

Vlassis, N. (2003), *A Concise Introduction to Multi-Agent Systems and Distributed AI, Intelligent Autonomous Systems*, Informatics Institute, University of Amsterdam, Holland.

Weiss, G., Ed. (1999). *Multiagent Systems: a Modern Approach to Distributed Artificial Intelligence*, MIT Press.

Information Exchange Requirements for Energy Audits in Commercial Building Retrofit Sector

Miaomiao NIU[1], Tabitha L. SPRAU COULTER[2], Robert M. LEICHT[3], A.M. ASCE
and Chimay J. ANUMBA[4], F.ASCE

[1] Graduate Research Assistant, Department of Architectural Engineering, Penn State, Engineering Unit A, University Park, PA 16802, mxn228@psu.edu
[2] Graduate Research Assistant, Department of Architectural Engineering, Penn State, Engineering Unit A, University Park, PA 16802, tls420@psu.edu
[3] Assistant Professor, Department of Architectural Engineering, Penn State, Engineering Unit A,   University Park, PA 16802, rmleicht@engr.psu.edu
[4] Professor and Head, Department of Architectural Engineering, Penn State, Engineering Unit A, University Park, PA 16802, anumba@engr.psu.edu

ABSTRACT

With an increasing demand for building energy performance, energy audits are being adopted more often by commercial building owners in order to determine the feasibility of building retrofits. There are many resources available for energy audits such as suggested data collection forms. However, there is not a standard that defines the information exchange for the energy auditing processes, which makes it difficult to leverage the information exchange to benefit downstream energy modeling. This paper describes the research on a standardized input data structure for energy audits in order to help define the energy auditing protocol for retrofit buildings. The research defines the standard required input data structure for Level II energy audits based on the content analysis of the current audit data templates and energy modeling input requirements. By aligning the data with the main energy modeling software tools, the consistent categories of data for energy audits are extracted. The benefits of the adopted approach are presented and future research trajectories are outlined.

INTRODUCTION

Improving the energy efficiency of commercial buildings has been recognized as a key strategy to reduce energy costs and to promote a sustained economic growth. As energy efficiency consciousness grows, energy audits are becoming a common practice in the construction industry (Zhu, 2005; Iqbal and Al-Homoud, 2007). Energy audits are generally considered as the first step of energy retrofit projects which paves the way for more detailed analysis and metering key system performance, along with defining the economic benefits specific to the facility. There are an increasing numbers of owners and facility managers turning to energy professionals to perform building energy assessments, especially in the commercial building sector. However,

the quality of the service and the implementation results vary (Sprau Coulter et al, 2012). There are challenges related to this process which create a cost and effort barrier to customer consideration of deep retrofit for buildings. For large commercial buildings, the problems can be even worse since the HVAC equipment and controls can be complex with airflows that may be unpredictable and uncontrolled (Shapiro, 2009).

Data collection and feasibility evaluation for energy retrofit projects is commonly referred to as an energy audit (ASHRAE, 2004). The data collected for energy analysis will significantly impact the analysis results (Zhu, 2005; Raham, et al., 2010). However, there is not a standard process for energy audit data collection. The whole data collection and analysis process can take several months for detailed data collection and analyses. Misinterpretation of the data, erroneous assumptions, and the collection of unnecessary information are common mistakes due to a lack of integrated and standardized process and data structure (Knapp, 2006). Moreover, current energy modeling software were developed primarily to simulate the energy loads of new buildings, where only comparative results are expected rather than for accurate and precise models of existing buildings (Waltz, 2000). Therefore, there is a need for an integrated and consistent data structure for energy audits in order to facilitate the information flow from energy auditing to energy analysis to improve the flow and reliability of data for existing facilities.

There are some public energy audit data collection sample forms that provide reference guide on the input data of energy audit. The *Procedures for Commercial Building Energy Audits* of American Society of Heating, Refrigerating and Air Conditioning Engineers (ASHRAE) establishes levels of audit effort and introduces procedures for energy auditors. The publication also provides sample data collection forms for energy audits (ASHRAE, 2004). The Rocky Mountain Institute (RMI) published another set of data collection template forms that assist building energy auditors to collect the data required for a comprehensive energy audit (RMI, 2011). However, these data collection forms are not aligned, which makes it difficult to identify the information to collect during the field data collection. Moreover, these sample forms only define the input information that is necessary for energy audits, the information flow from the energy audit to the energy analysis process is not considered. In order to address these issues, the currently available energy audit data guidelines and templates were analyzed to identify the complete and consistent categories. The categories were compared with the standardized data input of four common energy modeling tools in order to evaluate the potential leverage of the information exchange requirements from energy audit to energy modeling. The goals of this research include: 1) identifying a consistent data structure for an energy audit; 2) compare the standard audit data to energy modeling inputs for several commonly used energy modeling tools; and 3) identify the information exchange gaps and problems for common exchange from energy audit to energy modeling for retrofit projects.

AVAILABLE DATA STRUCTURES

Information Exchange Descriptions are forms that identify the information content of an exchange. They identify which objects, process properties, relations and classifications are both relevant to the receiving (importing) application and available in the sending (exporting) application. **Information Exchange Requirements** are specified in terms of the information items which carry the exchange descriptions (Eastman et. al, 2011). The purpose of defining information exchange requirements is to specify the information items in sufficient detail in order to facilitate the communication process. As previously discussed, there is no standard information exchange requirement for energy audits, which makes it complex and difficult to define and perform the process using standard auditing tools. There are some guidelines and templates available on the data suggested for commercial building energy audits, such as the ASHRAE publication and the RMI templates. The purpose of these data structures is to assist building energy auditors to collect data required for complete energy and financial analyses. The intent of the analysis is to identify gaps among these data structures which make it difficult to identify the right information.

Energy modeling is the focused step of analyzing the energy performance of buildings with the intent of evaluating architectural and/or mechanical designs through some energy simulation (Waltz, 2000). Energy simulation tools allow the design team to evaluate the impacts of various design options that are being considered, and to develop an effective building form and design strategies (GSA, 2009). Although more widely used for new building energy performance simulation and prediction, energy modeling also has opportunities for energy modeling in retrofit projects (GSA, 2009). Energy modeling provides accurate modeling of the real-time conditions and the existing building system energy performance evaluation. Various tools are used for energy modeling; the Department of Energy (DOE) identifies over 300 on their Building Energy Software Tools Directory (DOE 2013), with EnergyPlus, eQuest, Trane Trace 700 and Hourly Analysis Program (HAP) among the most common in industry. The relationship of energy audits and energy modeling from the perspective of information exchange is shown in Figure 1.

CONTENT ANALYSIS WORK FLOW

In order to integrate the energy audit and energy modeling information exchange requirements, the currently available energy data structures were analyzed and the relationships between these data structures were identified. For the energy audit templates, detailed data items were extracted from all sample forms and organized according to construction systems. Energy modeling input data was extracted manually from the software data entries.

The detailed data items under each data structure were cross compared, including the data description, data type and unit. Based on the analysis, the differing details by tool were addressed in order to extract the energy audit characteristics. The standardized data input of these energy modeling tools was analyzed in order to compare the data from the energy audit to establish the link between the energy audit and energy simulation process.
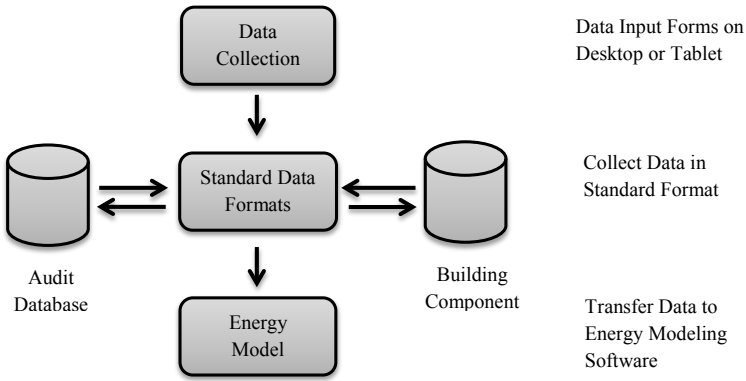
Figure 1. Data Flow between Energy Audit and Energy Modeling

RESULTS SUMMARY

Based on the workflow, the data item details from each available data structure was extracted from the template forms or software database and input into spreadsheets. Each item was cross compared to identify the overlaps and gaps between different data structures. Currently, ASHRAE's *Procedures for Commercial Building Energy Audits* is the only industry standard that provides guidelines for commercial building energy audits, which is comprehensive to cover the core data for different levels of energy audit. Therefore, the ASHRAE sample forms were referenced as the baseline for the analysis. Each data item was analyzed based on data description, data type and unit, and coded in spreadsheets. The data structures were cross compared against the ASHRAE standards based on the data item details. The overlapped data were marked and aggregated to calculate the percentage of data alignment. An example of the comparative details is shown in Table 1. The RMI data items in the building envelope category were compared against the ASHRAE data items. If the data item description conveyed identical information, the item was marked as an overlapped data. Items that convey no input value were excluded from comparison. The overlapped items were summarized by each category, in this example, ASHRAE and RMI had 8 overlapped data items out of 17 total items. Therefore, the data alignment is 47%.

Based on the analysis, the percentage of data alignment is summarized as shown in Table 2 and Figure 2. The core data was categorized into eight groups covered by all of the energy audit tools. Historical building energy use data in ASHRAE template forms was excluded from the analysis, as this category of information is not covered by most energy modeling software, though a notable gap when considering retrofit buildings and calibration processes.

Table 1. An Example of Cross Comparing Details between ASHRAE and RMI Data Structure

ASHRAE Data Collection Sample Forms			RMI Data Collection Sample Forms			Overlap	Total
Data Item	Type	Unit	Data Item	Type	Unit		
Total Exposed Above-Grade Wall Area	Real	ft²					1
Total Exposed Above-Grade Wall Area Insulated or Not? (Y/N)	Enum						0
Glazing Area (% of Exposed Wall Area)	Real	ft²	Glazing Dimensions (HxW)	Real	ft x ft	1	1
Glazing Area Single or Double?	Enum		Glazing Type	Enum		1	1
Roof Area	Real	ft²					1
Roof Insulated or Not? (Y/N)	Enum						0
Floor Surface Area Exposed to Outdoor Conditions	Real	ft²					1
Floor Surface Exposed to Outdoor Conditions Insulated or Not?	Enum						0
Above-Grade Wall Area Common With Other Conditioned Build	Real	ft²					1
Walls Construction Code	Enum		Layers Description	String		1	1
Doors Construction Code	Enum		Door Type	Enum		1	1
Roofs Construction Code	Enum		Layers Description	String		1	1
Windows Construction Code	Enum		Glazing Type; Frame Type	Enum		1	1
Walls R-Value	Real	hr-ft2-F/BTU	Walls Overall R-Value	Real	hr-ft2-F/BTU	1	1
Doors R-Value	Real	hr-ft2-F/BTU					1
Roofs R-Value	Real	hr-ft2-F/BTU	Roofs Overall R-Value	Real	hr-ft2-F/BTU	1	1
Walls Shading Coefficient	Real						1
Doors Shading Coefficient	Real						1
Roofs Shading Coefficient	Real						1
Windows Shading Coefficient	Real						1
						8	17
						47%	

From Table 2, it is fairly clear that there is relatively consistent emphasis on schedules, envelopes, and HVAC systems within the energy modeling tools. However, it is notable that there are systems entirely lacking in some of the tools, such as lighting system inputs, and limited use of domestic hot water systems. Also, with the shift in focus to energy efficiency in building systems, the need to focus on user loads and plug loads has grown tremendously in recent years with the impact on building performance which can be quite influential (Tutar, 2012). These findings suggest a strong need to further refine and identify standards for data exchange in energy data collection and modeling tool platforms.

CONCLUSION AND FUTURE WORK

Data collection on existing building conditions is significant to the detailed energy and economic analysis of energy retrofit projects. Identifying the information exchange requirements of energy audits offers potential to improve information flow from energy auditing to energy modeling processes by reducing the work of manual data entry. Mistaken data and re-collection could also be reduced. Furthermore, information exchange requirements may help to define the standardized process. However, there is no information exchange standard for energy audits and little research has been focused on energy audits in terms of information exchange requirements.

Table 2. Energy Data Alignment Summary

| Category | Energy Auditing | | | | | | Energy Modeling | | | | | | | | |
| | ASHRAE | | RMI | | eQuest | | Energy Plus | | HAP | | Trane Trace 700 | | | | |
	Items	Alignment	Items	Alignment	Items	Alignment	Items	Alignment	Items	Alignment	Items	Alignment
General Information	32	100%	11	34%	10	31%	9	28%	11	34%	6	19%
Utility	40	100%	23	58%	1	3%	3	8%	0	0%	2	5%
Schedules	5	100%	2	40%	2	40%	3	60%	3	60%	0	0%
Building Envelope	17	100%	8	47%	5	29%	9	53%	8	47%	4	24%
Lighting System	11	100%	7	64%	0	0%	2	18%	3	27%	0	0%
HVAC System	27	100%	10	37%	13	48%	14	52%	10	37%	11	41%
Domestic Hot Water	7	100%	4	57%	3	43%	4	57%	0	0%	0	0%
Special Loads & Other Systems	10	100%	3	30%	0	0%	0	0%	5	50%	0	0%
Total	149		68	46%	34	23%	44	30%	40	27%	23	15%
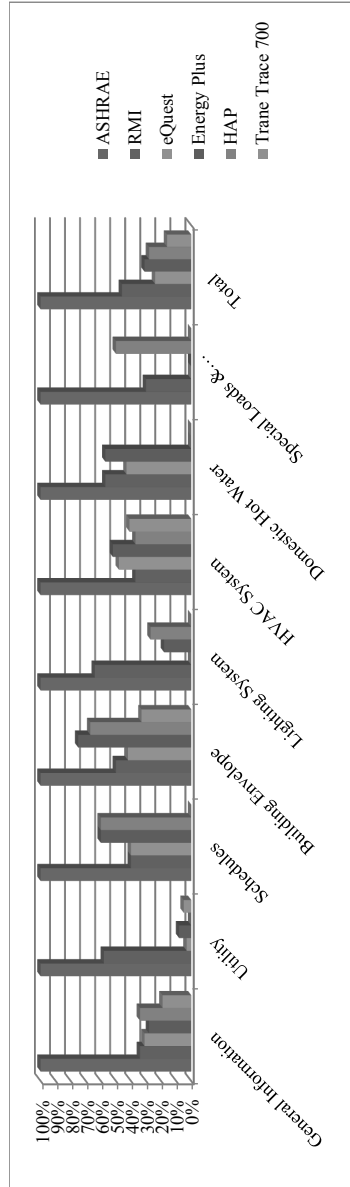
Figure 2. Energy Data Alignment Summary

This paper has described the information flow between energy auditing and energy modeling by analyzing the currently available energy audit template forms and energy modeling software input data for four common energy modeling applications. Based on the analysis results, the overall alignment of the data structure among current energy audit templates and energy modeling tools was determined to be low. The average alignment of the energy modeling inputs with ASHRAE template forms less than 25%. Three reasons contribute to the phenomenon: 1) most energy modeling tools are designed to simulate and predict the energy use of new buildings rather than existing buildings, thus there are several data input categories, such as utility bills, which are not standard inputs for energy performance modeling tools, 2) the level of detail collected is higher than typically used because the buildings are completely designed, and 3) the ASHRAE template forms are designed for both owners/facility managers and professionals to understand the procedures of energy auditing while most energy modeling software are designed for professionals only. In some cases, the applications may carry default values for system data, such as lighting densities, but are not specifically aligned to consume audit data for these systems. Also, ASHRAE identifies qualitative information for some system and equipment evaluation while most of the input information for energy modeling software tools is quantitative. It is essential as the focus on retrofit projects grow to leverage the audit data for detailed energy analysis. The identified overlaps and gaps should be analyzed in more detail in order to improve the detailed information flow from data collection work in energy audit in a more seamless fashion. Another benefit from analyzing the information exchange requirements of energy audits is to facilitate the field data collection process. For example, both ASHRAE and RMI consider the R-value of the building envelope system as an important index in energy audit. However, it is difficult to identify the R-value of the building system during data collection. While the energy modeling tools are using dimensions and materials, the R-value could be automatically calculated. Analyzing the differences is helpful to identify the energy audit characteristics in order to facilitate the data collection process.

This paper explored the information exchange requirements of energy audits in the commercial building retrofit sector. The results are beneficial to establishing the need for a standardized energy audit process model and information exchange. The adopted approach could be used to analyze the information exchange between processes with available data structures. Future research should focus on the validation of the current results with current energy audit walk-through analysis tools used for alignment with the identified standard data.

ACKNOWLEGEMENTS

REFERENCES

ASHRAE. (2004). *Procedures for Commercial Building Energy Audits*. Atlanta: American Society of Heating Refrigeration and Air Conditioning Engineers, Inc., RP-669, SP-56.

DOE. (2013). Building Energy Software Tools Directory, http://apps1.eere.energy.gov/buildings/tools_directory/alpha_list.cfm>(Jan. 1, 2013).

Eastman, C., Panushev, I., Sacks, R., Venugopal, M., Aram, S., See, R., Yagmur, E. (2011) A Guide for Development and Preparation of a National BIM Exchange Standard, http://www.buildingsmartalliance.org/client/assets/files/bsa/IDM-MVD_Development_Guide_v4.pdf>(Dec. 8, 2012).

GSA. (2009). "BIM Guide – Energy Performance and Operations." *United States General Services Administration*. http://www.gsa.gov/portal/content/105075>(Nov. 10, 2012).

Iqbal, I. and Al-Homoud M. (2007). "Parametric analysis of alternative energy conservation measures in an office building in hot and humid climate." *Energy and Environment*, 42, 2166–2177.

Knapp, A. (2006). "Lean Energy Audits: Lean Energy Auidts: Rethinking Common Management Practices of Multi-Building Energy Audits." *Strategic Planning for Energy and the Environment*, 25:4, 71-78.

Rahman, M., Rasul, M. and Khan, M. (2010). "Energy conservation measures in an institutional building in sub-tropical climate in Australia", *Applied Energy*, 87(10), 2994–3004.

RMI. (2011). *Energy Audit Sample Forms*. Rocky Mountain Institute. http://www.rmi.org/.../AuditSampleForms_Master_Sept_27_2011>(Nov. 10, 2012)

Shapiro, I. (2009). "Energy Audits in Large Commercial Office Buildings." *ASHRAE Journal*, 9:1,18-27.

Sprau Coulter, T., Hinsey, J., Leicht, R., Whelton, M., Riley, D. (2012). "Identifying energy auditing process and information exchange requirements for the commercial building retrofit sector." *Construction Research Conference*. West Lafayette, Indiana.

Tutar, T. (2012). Validating the impact of plug load management strategies on achieving deep energy retrofits. MS Thesis, The Architectural Engineering Department, Penn State, University Park, PA, 16802.

Waltz, J. (2000) "Computerized building energy simulation handbook." Liburn, Georgia.

Zhu, Y. (2005). "Applying Computer-Based Simulation to Energy Auditing: A Case Study." *Energy and Buildings*, 38(5), 421-428.

Link criticality based on most probable network states for pre-disaster investment

Kwangho Kim[1] and Yoonjin Yoon[2]

[1]Postdoctoral Researcher, Department of Civil and Environmental Engineering, University of California, Berkeley, CA, U.S.; PH 1-510-847-0066; FAX 1-510-642-0910; e-mail: khkim@berkeley.edu
[2]Assistant Professor, Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, South Korea; PH 82-42-350-3615; FAX 82-42-350-3610; e-mail: yoonjin@kaist.ac.kr

Abstract

In the context of pre-disaster investment, we propose an unconventional procedure to prioritize links by evaluating the impacts of their failures on the expected network performance. Our proposed procedure is termed as Link-based Categorization of Network States (LCNS) in the sense that network states are categorized according to each link's failure. In this unconventional method, links' criticality indices are cumulatively calculated, as most probable network states are sequentially enumerated. Our procedure is applicable to the situation when there is no prior knowledge concerning the effect of pre-disaster investment on the link survivability. Moreover, LCNS does not cause any redundancy in enumerating network states. An implementation of the proposed procedure on an idealized transportation network reveals that component links' ranking in terms of criticality indices is converged with only 0.2% enumeration of total network states plausible under a disaster.

1. Introduction

Man-made or natural disasters can significantly damage links constituting civil infrastructure networks. To mitigate system-wide impacts of network disruption, resources need to be optimally allocated in a proactive way. Prioritizing links in this context has been received attention of many researchers; see the review of Sullivan et al. (2010).

The survival probability of a link is determined by factors such as the link's current physical condition or retrofit history. The impacts of these factors on the link survivability are difficult to estimate in the real-world situation. Moreover, the problem of estimating the expected connectivity for a large network has been known as NP hard (Ball, 1979). For instance, if we consider binary conditions (0: failure, 1: operation) for each of n links, 2^n network states should be enumerated to exactly calculate the expected network performance.

The potential impact of a disaster on network performance would be affected by the investment policies implemented to individual links. In this context, one way of deriving optimal set of critical links by implementing stochastic mathematical programing techniques; see e.g. Peeta et al. (2010). This approach, however, is applicable only when we have accurate knowledge concerning the effect of investment policies on the survivability of individual links.

The present study is intended to determine link criticality indices by evaluating the impacts of links' failures on the performance of enumerated networks. We propose an unconventional method termed as Link-based Categorization of Network States (LCNS) to

cumulatively calculate links' criticality indices, as most probable network states are sequentially enumerated. This heuristic approach is applicable even without any prior knowledge about the effect of pre-disaster investment on the link survivability.

This paper is organized as follows. Section 2 summarizes relevant studies to determine the link criticality indices for a vulnerable network. Section 3 introduces an existing algorithm to enumerate most probable network states and illustrates how this algorithm can be used i) to estimate the upper and lower bounds of expected network performance and ii) to determine link criticality indices. Section 4 describes outcomes of implementing our proposed method on an idealized transportation network. Finally, we discuss practical implications of the proposed method in Section 5.

2. Literature Review

Some researchers attempted to find out *k*-most vital links, whose simultaneous removal results in the greatest increase in the cost of the shortest path connecting an origin-destination pair (see e.g. Malik et al., 1989). Identified vital links can be considered as candidates for pre-disaster investment. In line with this, Kurauchi et al. (2009) hypothesized removing individual links, one by one to determine the criticality index of each link by evaluating changes in the number of paths connecting each O-D pair within reasonable amounts of travel time. To further, Segovia et al. (2012) proposed a heuristic method to prioritize links of a network exposed to simultaneous failures of multiple links.

Taylor et al. (2006) argued that the network-wide consequence of a link's failure, irrespective of the link's failure probability should be a primary concern of pre-disaster infrastructure management for a sparsely connected regional transportation network. This argument is in line with the concept called 'network vulnerability,' which has been used to determine link criticality indices (see e.g., Sullivan et al., 2010; Jenelius and Mattsson, 2012).

Monte-Carlo simulation can be crudely applied to estimate the expected network performance by sampling independent network states, and then measuring the mean of sampled network performance values. This crude estimator shows a large variance in estimating the expected network performance, particularly when sample sizes are small; see Van Slyke and Frank (1972). Diverse techniques have been developed to improve the Crude Monte-Carlo approach (see e.g., Fishman, 1986; L'Ecuyer et. al, 2011). Further, Nojima (1998) prioritized links in terms of criticality indices obtained by evaluating the expected network performance via a variance-reducing Monte-Carlo technique.

Wakabayashi and Iida (1992) applied Boolean Algebra to calculate the upper and lower bounds of expected terminal connectivity for each origin-destination pair. Li and Silvester (1984) developed an algorithm to estimate bounds of expected network performance by enumerating most probable network states. Other researchers improved the efficiency of L&S's algorithm by employing data structures based on the graph theory (see e.g. Lam and Li, 1986; Yang and Kubat, 1989). Poorzahedy and Bushehri (2005) prioritized links by employing L&S's algorithm to evaluate the impacts of links' hypothetical removals on the expected network performance.

3. Methodology

We implement a tree search algorithm developed by Yang and Kubat (1989) to enumerate most probable network states. With enumerated network states, the upper and lower bounds of expected network performance are estimated using the formula proposed by Li and Silvester (1984). In addition, link criticality indices are determined by evaluating how the performance of each impaired network state (due to links' failures) degrades in comparison with the counterpart of the fully-connected network state.

3.1 Network model

We model a network whose links are prone to fail independently with one another. For simplicity, we assume that nodes are perfectly resilient to disasters. Under this basic assumption, we consider a directed network G= (V, E) with node set V and link set E, where each link is denoted by the index i (i=1, 2,…, n). Each link would either survive or fail in the face of a disaster. We assume that any intermediate conditions between survival and failure do not exist. Let p_i ($0 < p_i < 1$) denote the probability that link i survives. We use a Bernoulli random variable ξ_i to denote the condition of link i. That is, ξ_i =1, if link i survives under a disaster, and ξ_i=0, otherwise. The vector composed of all the Bernoulli random variables for the links belonging to E is denoted by X= (ξ_1 ξ_2 ξ_3 …ξ_n). We denote the kth network state and its attendant network performance by X^k and S(X^k), respectively (k=1, 2 , …, 2^n); and the realization probability of network state X^k by P(X^k). Then, the expected network performance can be exactly calculated by equation (1).

$$E[S(X)] = \sum_{k=1}^{2^n} S(X^k)P(X^k) \qquad (1)$$

Enumeration of all the possible states, however, is not computationally feasible for a large-scale network with large n. To address this combinatorial complexity, we attempt to estimate upper and lower bounds (denoted as S_u and S_L, respectively) of expected network performance. Let S_{MC} and S_{LC} denote values corresponding to the network performance for the most and least connected networks, respectively. Given m ($m \ll 2^n$) enumerated probable network states, S_u and S_L can be estimated by equations (2) and (3), respectively.

$$S_u = \sum_{k=1}^{m} S(X^k)P(X^k) + \{1 - \sum_{k=1}^{m} P(X^k)\}S_{MC} \qquad (2)$$

$$S_L = \sum_{k=1}^{m} S(X^k)P(X^k) + \{1 - \sum_{k=1}^{m} P(X^k)\}S_{LC} \qquad (3)$$

3.2 Estimation of expected network performance

This subsection presents how Y&K algorithm is used to estimate the expected network performance. For an illustration, we consider Network G composed of four failure-prone links, as shown in the left-hand side of Figure 1. Each link i (i=1, 2, 3, 4) spans length l_i and

independently fails with a probability q_i (=1-p_i). Information on p_i and l_i are shown next to the network. At each iteration stage, Y&K's algorithm identifies an address corresponding to a network state, which is enumerated by migrating from the root node to the bottom level in the ordered tree shown in Figure 1. For an example, (1 1 0 1) denotes the network state accompanying the failure of link 3, simultaneously with the survivals of all the other links. In this illustration, the total length of all the survived links is defined as the performance measure of a realized network. For an example, the performance of network state X^3= (1 0 1 1) is calculated as S(X^3)=30+0+20+45=95.
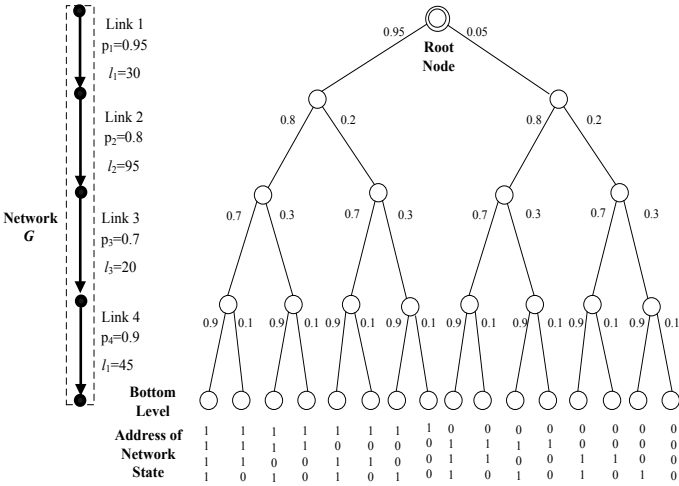
Figure 1 Enumeration of possible states of Network G

Table 1 presents outcomes of implementing Y&K's algorithm on Network G shown in Figure 1. The first column of this table shows the number corresponding to each iteration stage. In the second column, enumerated network states are listed. The third column represents the performance of each enumerated network. The fourth column shows the probability that each network state is realized. This column confirms that the algorithm enumerates network states in the decreasing order of realization probability. The fifth and sixth columns present upper and lower bounds of the expected network performance, respectively. Inspection of the values of the same row in these two columns reveals that at around the sixth iteration, the mean of S_U and S_L becomes close to the exact value of the network reliability (159 for the present case). The rightmost four columns show link criticality indices determined sequentially.

Table 1 Outcomes of implementing Y&K's algorithm on Network G

(1) Iteration Number k	(2) Network State X^k	(3) Network Performance $S(X^k)$	(4) Realization Probability $P(X^k)$	(5) Upper Bound S_U	(6) Lower Bound S_L	Link Criticality Indices			
						(7) $CI(L_1)$	(8) $CI(L_2)$	(9) $CI(L_3)$	(10) $CI(L_4)$
1	(1 1 1 1)	190	0.479	190	90.972	0	0	0	0
2	(1 1 0 1)	170	0.205	185.896	125.856	0	0	0.229	0
3	(1 0 1 1)	95	0.120	174.525	137.228	0	0.239	0.229	0
4	(1 1 1 0)	145	0.053	172.131	144.942	0	0.239	0.229	0.070
5	(1 0 0 1)	75	0.051	166.231	148.789	0	0.369	0.359	0.070
6	(0 1 1 1)	160	0.025	165.475	152.821	0.030	0.369	0.359	0.070
7	(1 1 0 0)	125	0.023	163.993	155.671	0.030	0.369	0.394	0.104
8	(1 0 1 0)	50	0.013	162.131	156.336	0.030	0.420	0.394	0.155
9	(0 1 0 1)	140	0.011	161.591	157.848	0.045	0.420	0.409	0.155
10	(0 0 1 1)	65	0.006	160.804	158.258	0.063	0.438	0.409	0.155
11	(1 0 0 0)	30	0.006	159.892	158.429	0.063	0.474	0.445	0.191
12	(0 1 1 0)	115	0.003	159.682	158.751	0.068	0.474	0.445	0.196
13	(0 0 0 1)	45	0.003	159.290	158.872	0.079	0.486	0.456	0.196
14	(0 1 0 0)	95	0.001	159.176	158.986	0.081	0.486	0.459	0.198
15	(0 0 1 0)	20	7.000E-04	159.057	159.000	0.088	0.492	0.459	0.205
16	(0 0 0 0)	0	3.000E-04	159	159	3.088	3.492	3.459	3.205

* Rows highlighted with shading and bold-boxes correspond to the network states associated with the failures of links 1 and 2, respectively.

3.3 Determination of link criticality using Link-based Categorization of Network States

LCNS cumulatively estimates link criticality by sequentially enumerating most probable states for Network G. This unconventional method categorizes network states according to the failure of individual links. For example, the category corresponding to the failure of link 1 includes network states such as (0 1 1 1), (0 1 1 0), (0 1 1 0), etc., as shown in the shaded rows in Table 1. Of course, this link-based categorization involves overlaps of different categories. For example, a realized network can involve failures of links 1 and 2 at the same time; note the simultaneous occurrences of shading and bold-box in Table 1. These categorical overlaps, however, are intrinsic characteristic of network realization, not the cause of redundancy in enumerating network states.

LCNS updates the critical indices of links at Y&K algorithm's each iteration stage. The detailed procedure is: whenever link i turns out to fail at an iteration stage, the link's critical index is increased by the product of 'probability of network realization' and 'degradation in the network performance' at the corresponding iteration stage. For example, at the sixth iteration, only link 1 failed, simultaneously with the other links' survivals; see the corresponding row in Table 1. In this case, the increment for link 1's criticality index is calculated as 0.025*(190/160)=0.030 while the other links' criticality indices remain the same; confirm this by comparing rightmost four values in the fifth row and their counterparts in the sixth row. In this way, the critical indices of four links are accumulated as the iteration number goes up. Since the eighth iteration, the ranking of four links in terms of criticality remain the same; i.e., $CI(L_2)> CI(L_3)> CI(L_4) > CI(L_1)$, for the present case.

4. Experiment

LCNS is tested using an idealized road network shown in Figure 2. Survival probability (p_i) and travel cost (C_i) associated with each link are shown in the table next to the network. We consider a situation when people should quickly evacuate from node 1 using the shortest-path connecting each of three origin-destination pairs (1→8, 1→9, and 1→10) in the test network.

Link#	P_i	C_i
1	0.989	2
2	0.988	5
3	0.977	11
4	0.989	5
5	0.954	13
6	0.654	5
7	0.957	32
8	0.955	16
9	0.963	6
10	0.700	2
11	0.675	3
12	0.867	6
13	0.832	15
14	0.936	12
15	0.993	13
16	0.833	7
17	0.937	10
18	0.799	9
19	0.762	7
20	0.800	29
21	0.763	16
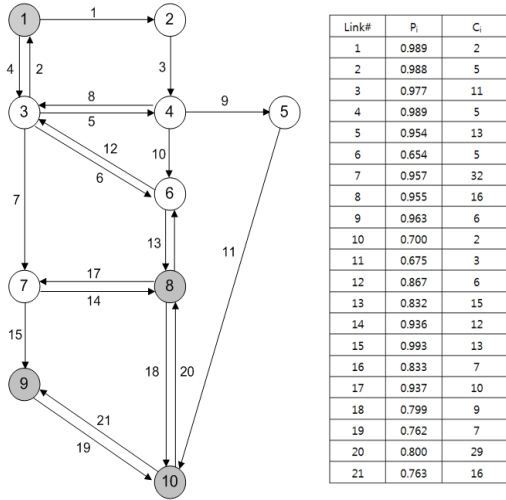
Figure 2 Test network and link information

The network performance here is defined as the terminal connectivity between each of the three O-D pairs. For simplicity, we assume that the terminal connectivity for an O-D pair is reversely proportional to the minimum travel cost required for the O-D pair. In the case that any connected path is not available for an O-D pair of a realized network, a predefined large penalty value is assigned to the minimum O-D travel cost. The degradation in the network performance is defined as the ratio between the minimum O-D travel cost determined for a realized network and its counterpart of the fully connected network.

Table 3 presents outcomes of implementing LCNS on the test network shown in Figure 2. For brevity, we just present two set of indices and ranks (determined in terms of the link criticality) obtained through implementing 4000 and 5000 iteration stages, respectively, for each O-D pair. The table confirms that each link's rank converges after sufficient iteration; note the consistency in each link's rank across the two iteration stages for each O-D pair. The cumulative probabilities of network states enumerated up to these two iteration stages amount to 0.943 and 0.954, respectively. Notably, only about 0.2% of total network states have been enumerated to achieve these cumulative probabilities. For all the three O-D pairs, links 6, 10, and 11 belong to the highest five ranks in terms of criticality indices; see the shading of the rows corresponding to these three links.

Table 3 Outcomes of implementing LCNS on the test network

Link Index i	O-D (1→8)				O-D (1→9)				O-D (1→10)			
	N_i=4000		N_i=5000		N_i=4000		N_i=5000		N_i=4000		N_i=5000	
	CI(L_i)	Rank	CI(L_i)	Rank	CI(L_i)	Rank	CI(L_i)	Rank	CI(L_i)	Rank	CI(L_i)	Rank
1	0.008	19	0.009	19	0.007	20	0.008	20	0.010	19	0.011	19
2	0.011	18	0.013	18	0.009	19	0.010	19	0.011	18	0.013	18
3	0.020	17	0.022	17	0.019	17	0.020	18	0.027	17	0.030	17
4	0.008	20	0.009	20	0.007	21	0.008	21	0.009	20	0.010	20
5	0.043	14	0.047	14	0.039	14	0.042	14	0.057	15	0.063	15
6	0.594	1	0.613	1	0.407	2	0.417	2	0.698	4	0.724	4
7	0.144	12	0.157	12	0.094	11	0.103	11	0.139	11	0.152	11
8	0.042	15	0.045	15	0.038	15	0.041	15	0.056	16	0.061	16
9	0.036	16	0.038	16	0.036	16	0.038	16	0.089	13	0.098	13
10	0.497	3	0.516	3	0.348	3	0.358	3	0.594	5	0.617	5
11	0.499	2	0.518	2	0.443	1	0.457	1	1.245	1	1.290	1
12	0.157	10	0.164	11	0.135	10	0.140	10	0.215	10	0.228	10
13	0.409	4	0.431	4	0.205	8	0.213	8	0.370	7	0.390	7
14	0.154	11	0.167	10	0.059	13	0.062	13	0.111	12	0.120	12
15	0.004	21	0.004	21	0.019	18	0.022	17	0.007	21	0.009	21
16	0.202	9	0.213	9	0.173	9	0.180	9	0.280	9	0.295	9
17	0.064	13	0.069	13	0.062	12	0.067	12	0.087	14	0.095	14
18	0.252	8	0.263	8	0.216	6	0.223	6	0.852	3	0.888	3
19	0.323	6	0.337	6	0.259	5	0.266	5	0.889	2	0.926	2
20	0.329	5	0.347	5	0.215	7	0.222	7	0.347	8	0.363	8
21	0.305	7	0.317	7	0.307	4	0.318	4	0.424	6	0.441	6

* Shaded numbers correspond to the outcomes whose corresponding links are ranked as one of the highest five under each scenario defined by both the O-D pair and the number of iteration (N_i).

5. Discussions

An experiment reveals that LCNS can produce a converged ranking of links in terms of criticality indices with only 0.2% enumeration of total states feasible for an idealized transportation network. For this experiment, we defined the network performance as the terminal connectivity without considering links' capacities or congestion effects. To further, we plan to test more general applications of LCNS on real-world transportation networks. For example, the network performance can be refined to incorporate travelers' route choice behaviors in response to congested links.

The caveat of LCNS is that we need to figure out the iteration stage, at which the rank of critical links starts to converge. In this regard, further research is required to systematically determine an optimal number of iteration stages. In addition, we need to figure out how to efficiently store relevant information in the process of recording enumerated network states, particularly for large-scale networks.

References

Ball, M. O. (1979). "Computing network reliability." *Operations Research*, 27(4), 823-838.

Fishman, G. S. (1986). "A Monte Carlo sampling plan for estimating network reliability." *Operations Research*, 34(4), 581-594.

Jenelius, E., and Mattson, L. (2012). "Road network vulnerability analysis of area-covering disruptions: A grid-based approach with case study." *Transportation Research Part A*, 46(5), 746-760.

Kurauchi, F., Uno, N., Sumalee, A., and Seto, Y. (2009). "Network evaluation based on connectivity vulnerability." *Transportation and Traffic Theory*, 637-649.

Lam, Y. F., and Li, V. O. K. (1986). "An improved algorithm for performance analysis of networks with unreliable components." *IEEE TRANSACTIONS ON COMMUNICATIONS*, 34(5), 496-497.

L'Ecuyer, P., Rubino, G., Saggadi, S., and Tuffin, B. (2011). "Approximate zero-variance importance sampling for static network reliability estimation." *IEEE TRANSACTIONS ON RELIABILITY*, 60(30), 590-604.

Li, V. O. K., and Silvester, J. A. (1984). "Performance Analysis of networks with unreliable components." *IEEE TRANSACTIONS ON COMMUNICATIONS*, 32(10), 1105-1110.

Malik, K., Mittal, A. K., and Gupta, S. K. (1989). "The k most vital arcs in the shortest path problem." *Operations Research Letters*, 8, 223-227.

Norjima, N. (1998). "Performance-based prioritization for upgrading seismic reliability of a transportation network." *Journal of Natural Disaster Science*, 20(2), 57-66.

Peeta, S., Salman, F. S., Gunnec, D., and Viswanath, K. (2010). "Pre-disaster investment decisions for strengthening a highway network." *Computers and Operations Research*, 37(10), 1708-1719.

Poorzahedy, H., and Bushehri, S. (2005). "Network performance improvement under stochastic events with long-term effects." *Transportation* 32, 65-85.

Segovia, J., Vilà, P., Calle, E., and Marzo, J. L. (2012). "Improving the resilience of transport networks to large-scale failures." *Journal of Networks*, 7(1), 63-72.

Sullivan, J. L., Novak, D. C., Aultman-Hall, L., and Scott, D. M. (2010). "Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: a link-based capacity-reduction approach." *Transportation Research Part A*, 44, 323-336.

Taylor, M. A. P., Sekhar, S. V. C., and D'Este, G. M. (2006). "Application of accessibility based methods for vulnerability analysis of strategic road networks." *Network and Spatial Economics*, 6, 267-291.

Van Slyke, R., and Frank, H. (1972). "Network Reliability Analysis: Part I." *Networks*, 1, 279-290.

Wakabayashi, H., and Iida, Y. (1992). "Upper and lower bounds of terminal reliability of road networks: an efficient method with Boolean Algebra." *Journal of Natural Disaster Science*, 14(1), 29-44.

Yang, C., and Kubat, P. (1989). "Efficient computation of most probably states for communication networks with multimode components." *IEEE Transactions on Communications*, 37(5), 535-538.

Demand Response in Buildings:
Engaging Thermostatically Controlled Loads in the Power Grid

E. C. Kara[1] and M. Bergés[2]

[1] PhD Student, Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890; e-mail: eckara@cmu.edu
[2] Assistant Professor, Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890; e-mail: marioberges@cmu.edu

ABSTRACT

Buildings accounted for nearly 75% of the electricity use in 2010, the largest portion among all sectors in the United States. Despite their contribution to the overall electricity demand, they have traditionally been considered as passive end-users of energy, and research in building engineering has largely focused on solutions to curtail their energy usage and improve energy efficiency. However, recently, the potential of buildings to become active participants in the electricity grid by providing ancillary services via direct load control has been garnering interest in the research community. In this paper, we introduce different demand response programs that use thermostatically controlled loads (TCLs) available in buildings. Specifically, we shed light on the existing work on direct load control for TCLs and identify the upcoming challenges associated with this approach. Finally, we introduce *BUFFER*: the building frequency forecast and electricity regulation framework, a novel decentralized and autonomous framework that uses TCLs in buildings to do frequency regulation for the power grid.

INTRODUCTION

From a traditional power architecture perspective, buildings are major electricity consumers in today's power grid. As more consuming sectors grow dependant on digital devices, consumers are asking more from the power grid in terms of both capacity and reliability, and tolerance of power quality drops such as brownouts and blackouts has fallen (Gellings et al. 2009). However, the increasing demand for high-quality power is still being met by inefficient infrastructure, and imbalances in demand and supply often yield blackouts and/or brownouts. Per year, U.S. customers are likely to experience 1.5 to 2 power interruptions and spend between 2 and 8 hours without power (Kassakian and Schmalensee 2011).

Until now, buildings have had no active role in satisfying customers' expectations regarding power quality and availability. Services employed to maintain power reliability (i.e., ancillary services) and regulations of frequency and voltage are typically provided by fossil fuel power plants, which remain idle for the majority of

the time and then react during short periods of peak demand to provide power. For example, a study of New York and New England showed that for more than 88% of the time, the region was using less than 30% of its generation capacity (Kassakian and Schmalensee 2011).

Recent studies show that buildings can provide power regulation and ancillary services through adjusting their demand to lower electricity consumption when the system reliability is imperiled (Callaway 2011, Faruqi et al. 2009, Kiliccote et al. 2011). Demand response (DR), the concept under which this idea falls, has been implemented in the U.S. since the 1970s. Readers may be familiar with the use of utility-controlled thermostats that utilities offer to customers along with monetary incentives to reduce peak load on hot summer days. Although FERC's report does not take into account autonomous, decentralized control strategies, it shows that DR could reduce the forecasted peak demand for 2017 by 14% (Faruqi et al. 2009).

In 2010, buildings (which span all sectors) accounted for nearly 75% of U.S. electricity consumption. Thermostatically controlled loads (TCLs), which include heating, ventilation, and air conditioning (HVAC) systems as well as refrigeration equipment, constituted more than 50% of the loads for most of the buildings (Department of Energy 2012). Thus, there is great potential for buildings to be engaged in demand response programs.

As expected, most research in demand-side services involving buildings has focused on using TCLs to provide load shedding during short periods of peak demand. This is typically accomplished with preprogrammed control equipment. Recently, a need has emerged to use TCLs for fast (close to real-time) demand response services using internet-based signals. For these types of DR mechanisms to be accurate and reliable, a rethinking of the existing DR programs is necessary, and the expertise of facility managers, building engineers and other domain experts will be crucial.

In this paper, we shed light on different DR programs that use TCLs, discuss the need for direct load control (DLC), and identify the upcoming challenges associated with this approach. Finally, we introduce *BUFFER*: the building frequency forecast and regulation framework, a novel decentralized and autonomous framework that uses buildings to do frequency regulation for the power grid.

DEMAND RESPONSE PROGRAMS

DR programs fall into two general categories: pricing programs and incentive-based programs. In a pricing program, utilities prompt customers to modify their energy use by adjusting electricity rates throughout the day (Newsham and Bowker 2010), thus motivating load shedding and peak reduction. Rather than operating under a flat-rate structure, pricing programs reflect the variations in the cost of power transmission based on when the power is being used (Newsham and Bowker 2010). Most pricing programs employ time-of-use pricing, real-time pricing, critical peak pricing, or peak-time rebates. In pricing program strategies, customers are informed of the electricity price on a day-ahead to an hourly basis. We suggest the following works on pricing programs for the interested reader: Newsham and Bowker 2010; Kassakian and Schmalensee 2011.

Incentive-based programs typically use baselines to estimate the benefits of DR action, and customers receive payments based on these benefits. Major incentive-

based programs are direct load control (DLC) and market-based programs. DLC strategies use special equipment to control demand while maintaining constraints agreed upon by the customer. Market-based programs use a bidding schema to contribute to DR services. Customers bid on energy and capacity markets, offering price and quantity pairs for total demand, and the optimal power flow (OPF) constraints are created accordingly. The OPF problem is then solved via an increased number of loads and constraints in the power system. To keep the OPF constraints manageable, researchers envision an additional layer of aggregators between the power markets and the customers who are participating via smaller loads. In the U.S., these aggregators will likely be private companies, retailers or the utilities themselves. They will quantify customer's DR benefits and provide incentives based on those benefits (Mathieu et al. 2012a).

Emerging DR and Challenges

Most of the recent DR programs in the U.S. focus on seasonal peak load reduction for commercial and industrial buildings. These programs use preprogrammed peak load shedding strategies, employing centralized open-loop control activated manually by facility managers or automatically by existing energy management and control systems (EMCS). Often, the DR signal is received on a day-ahead to hour-ahead basis. These types of DR mechanisms are classified as *slow* based on their response time. Recent work has shown an emerging need for *fast* DR services (i.e., ancillary services) capable of providing automated closed-loop control for load shedding and following in real time (Kiliccote et al. 2011). This can be achieved by engaging individual loads as part of small commercial and residential buildings, as well as investigating methods to increase the responsiveness of large commercial and industrial buildings to demand response signals.

DLC THROUGH TCLS

Moving towards *fast* DR services while keeping the power system reliable and the OPF problem tractable is challenging when pricing and market-based programs are used. Although customers receive instantaneous pricing signals, an optimization problem needs to be solved on the customers' side and responses to the signals are often unpredictable (Mathieu et al. 2012a). More predictable response can be obtained via market-based programs since the OPF problem is solved according to the information received from all bidders. However, an increase in the number of participants multiplies the number of constraints in the optimization, often making the problem unsolvable. We therefore conclude that these programs are more suitable for providing *slow* DR services. By their nature, DLC programs can provide a faster response since individual loads are often automated and controlled directly based on grid instability information (Mathieu et al. 2012a).

As emphasized above, TCLs constitute a significant portion (more than 50%) of total U.S. electricity consumption, and are widely available in most U.S. households. It is possible to toggle them ON/OFF or control their power usage without significantly influencing their end-use function, and they can be easily modeled using simple heat transfer equations (Callaway and Hiskens 2011; Callaway

2009). All of these properties make TCLs favorable for a dynamic control mechanism capable of providing a faster response via DLC.

In the following sections, we introduce the requirements for using TCLs as DLC agents, as well as suggest solutions to fulfill these requirements and present initial results assessing these solutions.

Challenges and Requirements

To address the need for a faster DR service, research in DLC through TCLs has focused on closed-loop and centralized control strategies. Figure 1 shows a simplified DLC schema that uses a similar strategy. As described previously, the aggregator layer can be a private company, a retailer or the utility. It controls a population of the loads using information received from the utility on current grid conditions. The *downstream* interactions between the utility and the loads are provided by DR signals that carry information on grid instability and control commands for loads. The *upstream* interactions are typically load-dependent and can include both real-time load consumption data and historical information on the load to allow for load profile predictions (Koch and Piette 2009).

Other than real-time power consumption, recent studies focusing on TCLs to provide DLC services emphasize the need for additional upstream information to provide reliable power control while keeping the end use uninterrupted (Callaway et al. 2011; Mathieu et al. 2012b; Kara et al. 2012). These studies require additional one-time information that must be sent to the aggregator. This one-time information includes the thermal features and rated power value of TCLs, and is needed to create accurate models of the population on the aggregator side for non-disruptive control of TCLs. Figure 1 shows an exemplary population of TCLs formed by a water heater and a refrigerator. Each TCL should be aware of its thermal features and rated power consumption, and be able to send this information to the aggregator, who then uses this to create a population model to predict and control the aggregate demand.
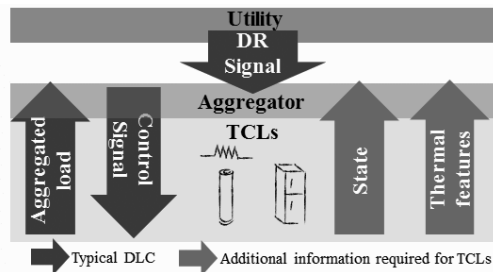
Figure 1. Centralized DLC Schema and Information Requirements.

In case of significant changes in the TCL population, the aggregator should update the population model using similar information received from the TCLs. For example, when a new HVAC system is added to the population, the aggregator should receive the information from the new load and update the model accordingly. Furthermore, in an ideal scenario, the aggregator needs to receive state information

from each individual load close to real time to provide reliable control over the loads. This information is formed by the status of the appliance (i.e. ON or OFF) and its current temperature with respect to the temperature deadband of the thermostat.

The increase in communication requirements of the ideal scenario prompted researchers to investigate the trade-off between the amount of required information and control accuracy. A recent study used state estimation techniques to infer the current condition of the population without receiving the state information from TCLs (Mathieu et al. 2012b). Instead, aggregate power consumption of the population is assumed to be available to the aggregator to estimate individual states of the loads.

Hence, recent studies of centralized DLC strategies introduce certain privacy concerns: *these strategies require that TCLs should be aware of their thermal features, rated power consumption, and real-time state information, and that they should be transferable via upstream information exchange between the loads and the aggregator depending on the control accuracy required*. In the following sections, we introduce a decentralized and autonomous framework to address these challenges and shed light on needed future research.

Building Frequency Forecast and Electricity Regulation (BUFFER) Framework

A power system's frequency is the primary indicator of its supply and demand imbalance. In the U.S., the system frequency is 60Hz under ideal conditions. It decreases when there is lack of generation and increases when there is a generation excess. Furthermore, it is observable anywhere in the power network, making it a great replacement for a DR signal. The BUFFER framework uses these fundamental principles and a Model Predictive Control strategy to turn buildings into autonomous and decentralized agents that provide frequency regulation.

Figure 2 depicts the BUFFER framework. The forecasting layer uses any socket available in the building to accurately sample the system frequency and predict the system frequency for a short horizon in real time. The energy management and control layer then uses a model predictive control (MPC) framework to minimize the deviations of frequency by looking at the forecast values and the TCLs' power consumption. Note that the flexibility of an MPC framework enables us to include constraints on user preferences in the optimization routine.
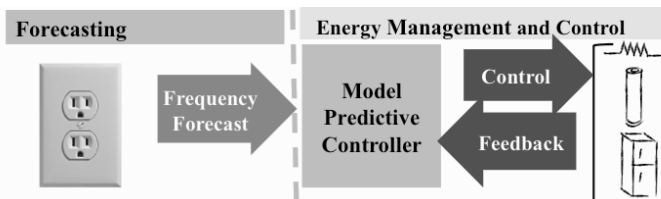
Figure 2. BUFFER Framework.

The energy management and control layer uses feedback obtained from the TCLs, which includes state information and rated power consumption, similar to the

centralized DLC scenario. However, this framework does not require information exchange between the aggregator and the loads. This will *ensure* privacy. For buildings with an EMCS (i.e., large commercial buildings), state information from controlled TCLs is often accessible. For customers interested in using the BUFFER framework for plug-level TCLs (i.e. a refrigerator or water heater), obtaining the state information can be challenging. Today's appliances cannot provide this information to the customer, and in a typical usage case, the customer will manually interact with the appliance via its own thermostat and adjust the settings according to his or her needs. An ideal solution would not tamper with the appliance to obtain the state information and would employ the BUFFER framework via a plug-and-play strategy.

Preliminary Case Study

To investigate the feasibility of the BUFFER framework and to analyze ways to provide real-time state information for the centralized controller developed in our previous work (Kara et al. 2012), we performed preliminary studies by collecting power consumption information from residential refrigerators. Figure 3 shows a wireless sensor node that was used to implement the functionalities of the framework and obtain the needed contextual information. This particular sensor node is based on the FireFly platform (Rowe et al. 2009) and can sample current and voltage at the plug to estimate power consumption and frequency.

Power consumption of TCLs can be a clear indicator of the state that the load is in at any given time. Hence, we wanted to investigate the feasibility of obtaining the state information from real-time power readings using currently available technology. The left plot in Figure 3 shows the power consumed by one of the refrigerator units without environmental changes or user interaction (e.g., opening/closing the door). The periodic nature of this time-series is due to the cycling of the compressor to maintain a specific thermostatic set-point. The start of each cycle matches the upper bound of the thermostatic dead-band, while the end of each cycle corresponds to the lower bound. The length of each cycle is related to the thermal properties of the refrigerator, as well as the user interventions and environmental conditions.
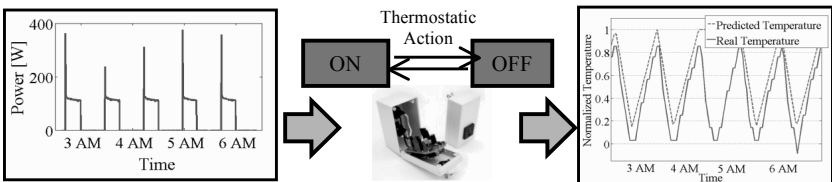
Figure 3. Sensor/Actuator Box and State Prediction Model.

To detect the start and end of each cycle from power measurements alone, we first employed a k-means clustering algorithm to group the power readings into 2 different bins corresponding to refrigerator ON and refrigerator OFF states. We then used the transition between these bins to detect a status change and identify the direction of the transition. To test this algorithm, we have collected approximately 80

hours worth of data from three different refrigerators, making sure that there were no disturbances to the refrigerators during this time. We ran the algorithm 11 times with randomly selected starting points from the data. Following that, we built a confusion matrix, and calculated precision and recall values for detecting state changes and correctly predicting their direction (i.e., turning on or turning off). Table 1 shows the results.

The right plot in Figure 3 shows the results of the state prediction model of a freezer, based on the clustering algorithm described above, in comparison to the ground truth. The ground-truth values are normalized so that the temperature lower-bound and the upper-bound of the TCL are set to 0 and 1 respectively.

The results shown in Table 1 indicate that in the absence of environmental changes or user interactions, detecting the state changes and obtaining the state information is trivial. It remains to be seen whether doing so while taking into account possible changes in the thermal properties, or user interactions will affect these results. However, we believe that there is an opportunity to incorporate this information into the BUFFER framework by learning these patterns from historical data collected over time.

	Average Precision	Average Recall
State Change Detection	97.61%	100.00%
OFF ➔ ON	97.06%	98.02%
ON ➔ OFF	98.04%	97.09%

Table 1. State Change Detection Results

CONCLUSIONS AND FUTURE WORK

This paper introduced the existing work on using TCLs in buildings as DLC agents for DR services and identified the upcoming challenges associated with this approach. We introduced a novel decentralized framework that uses TCLs in buildings to conduct frequency regulation, *BUFFER*, and gave our initial ideas on how to retrieve the state information from power measurements obtained through plug-level meters connected to TCLs. This information provides valuable feedback to the controller that enables us to further investigate the limits of our work, while ensuring the users' comfort. Future research topics include: exploring frequency forecasting techniques, extending the user context and state prediction framework, and investigating model predictive control approaches that use forecasted frequency information to provide frequency regulation via building loads for the grid.

ACKNOWLEDGMENTS

REFERENCES

Callaway, D. (2009). "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy." *Energy Conversion and Management*, 50(5), 1389–1400.

Callaway, D. (2011). "Can smaller loads be profitably engaged in power system services?." *Power and Energy Society General Meeting, 2011 IEEE*, 1 –3 (July).

Callaway, D., Koch, S., and Mathieu, J. (2011). "Modelling of control of aggregated heterogenous thermostatically controlled loads for ancillary services." *Power Systems Computation Conference*.

Callaway, D. S. and Hiskens, I. A. (2011). "Achieving controllability of electric loads." *Proceedings of the IEEE*, 99(1), 184–199.

Department of Energy, U. (2012). "2011 Buildings Energy Data Book.

Faruqi, A. H. R., S.S., G., Bode, J., Mangasarian, P., Rohmund, I., Wikler, G., and Glosh, D. Yoshida, S. (2009). "A national assesment of demand re- sponse potential." *Report no.*, Federal Energy Regulatory Commission (FERC), <www.ferc.gov/legal/staff-reports/06-09-demand-response.pdf> (June).

Gellings, C., Samotyj, M., and Howe, B. (2004). "The future's smart delivery system [electric power supply]." *Power and Energy Magazine, IEEE*, 2(5), 40 – 48.

Kara, E. C., Berges, M., Krogh, B., and Kar, S. (2012). "Using smart devices for system-level management and control in the smart grid: A reinforcement learning framework." *IEEE SmartGridComm'12*.

Kassakian, J. and Schmalensee, R. (2011). "The future of the electric grid: An interdisciplinary mit study." *Report no.*, Technical report, Massachusetts Institute of Technology.

Kiliccote, S., Piette, M. A., and Ghatikar, G. (2011). "Smart buildings and demand response." *AIP Conference Proceedings*, 1401(1), 328–338.

Koch, E. and Piette, M. (2009). "Direct versus facility centric load control for automated demand response." *Grid Interop*.

Mathieu, J. L., Haring, T., and Andersson, G. (2012a). "Harnessing residential loads for demand response: Engineering and economic considerations." University of Florida, Gainesville (December).

Mathieu, J. L., Koch, S., and Callaway, D. S. (2012b). "State estimation and control of electric loads to manage real-time energy imbalance." *Power Systems, IEEE Transactions on*, PP(99), 1.

Newsham, G. R. and Bowker, B. G. (2010). "The effect of utility time-varying pricing and load control strategies on residential summer peak electricity use: A review." *Energy Policy*, 38(7), 3289–3296.

Rowe, A., Berges, M., Bhatia, G., Goldman, E., Rajkumar, R., Soibelman, L., Garrett, J., and Moura, J. (2009). "Demonstrating Sensor Andrew: Large-Scale Campus-Wide Sensing and Actuation." *Demo Abstract, IPSN*.

A Numerical DAE Approach for Solving a System Dynamics Problem

A. Shadpour[1], A.J.A Unger[2], M.A. Knight[3], C.A. Haas[4]

[1]Department of Earth and Environmental Sciences, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1; PH(519) 888-4567; email: ashadpou@uwaterloo.ca
[2]Department of Earth and Environmental Sciences, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1; PH(519) 888-4567; email: aunger@uwaterloo.ca
[3]Department of Civil and Environmental Engineering, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1; PH(519) 888-4567; email: maknight@uwaterloo.ca
[4]Department of Civil and Environmental Engineering, University of Waterloo, 200 University Ave. W., Waterloo, Ontario, Canada N2L 3G1; PH(519) 888-4567; email: chaas@uwaterloo.ca

ABSTRACT

A system dynamics model first developed using the software STELLA, which explores the complex behaviour of the financially sustainable management of wastewater distribution infrastructure, was converted here into a system of coupled non-linear differential algebraic equations (DAEs). Each differential equation involved a time derivative on a primary variable specifying the temporal evolution of the system. In addition, algebraic (secondary) equations and variables specified the non-linearity inherent in the system as well as any controls on the primary variables constraining the physical evolution of the system relevant to the problem at hand. While STELLA employed a Runge-Kutta numerical strategy, the numerical DAE method used a fully-explicit scheme combined with a fixed-point iteration to resolve the non-linearity. The Runge-Kutta and numerical DAE solutions deviate markedly when the non-linearity of the system becomes pronounced. We demonstrate point-wise stability of the numerical DAE solution as the timestep is refined.

INTRODUCTION

System dynamics evolved in the 1950's (Forrester, 1958) as a computation tool to quantify the behaviour of complex transient systems. The mathematical vernacular associated with system dynamics involves breaking a problem down into stocks, flows, converters and connectors. Stocks are used to capture the transient behaviour of select variables within the problem using discrete time derivatives. Flows, convertors and connectors are used to specify mathematical relationships needed to transmit information between the various stocks.

Sepplet and Richter (2005) and Rizzo et al. (2006) have called into question the numerical accuracy of these various software packages. Rehan et al. (2011) present simulation results using STELLA 7.0.2 for which some variables exhibit what

appear to be spurious oscillations; otherwise, local maxima or minima which may indicate non-monotone behavior of the solution. These same spurious oscillations appear in Son et al. (2011) and, in the case of their work, are interpreted as productivity dynamics within the context of a construction management system.

This work centres around the hypothesis that the set of equations developed using system dynamics vernacular can be translated directly into a series of coupled and potentially non-linear differential equations as well as a series of algebraic equations, commonly termed "differential algebraic equations" or DAEs (Petzold, 1982; Brenan et al., 1989; Ascher and Petzold, 1998). The premise is that DAE solution will be absent any spurious oscillations and be physically correct for the engineering problem at hand.

A NUMERICAL DAE APPROACH

The objective of this work is to specifically focus on the wastewater model of Rehan et al. (2013a, 2013b) as a prototype problem to highlight the proposed approach for converting a system dynamics problem into a series of coupled non-linear ordinary differential equations. Rehan et al. (2013a, 2013b) present the entirety of the equations constituting the system dynamics problem in the STELLA 7.0.2 notation of "stocks", "flows", "connectors" and "convertors". This same system is presented graphically on Figures 7, 8 and 9 of Rehan et al., (2013a). The premise in this section is that the stocks in system dynamics can instead be called primary variables within a set of DAEs, with the flows, connectors and convertors constituting the remainder of a given DAE as well as secondary equations and variables. To begin, let X denote the vector of primary variables identified from a given system dynamics problem and $Y(X)$ be the vector of secondary variables, some of which may be non-linearly dependant on the primary variables. The general form of a semi-explicit index-1 DAE system is given by Brenan et al. (1989):

$$\frac{\partial X}{\partial t} = f(X, Y, t)$$
$$0 = g(X, Y, t) \tag{1}$$

The notion of semi-explicit DAE (and explicitness in general) implies that the algebraic equations are separated from the differential equations as presented in Equation (1) above.

Brennan et al. (1989) review the application of Runge-Kutta and Backward Differentiating Formulations (i.e. Euler methods) for the numerical solution of DAEs. For semi-explicit index-1 DAEs, they demonstrate that the use of an Euler method of order $k \leq 7$ evaluated with a constant time step size Δt converges with an accuracy of order $\mathcal{O}(\Delta t^k)$ after $k + 1$ fixed-point iterations provided each fixed point iteration is solved to an accuracy of order $\mathcal{O}(\Delta t^{k+1})$. In other words, as the timestep Δt is progressively reduced along with the numerical error in approximating the time derivative, the solution should converge to a single solution over the entire simulation period and thereby exhibit point-wise stability. For semi-explicit index-1 DAEs, Bendtsen and Thomsen (1999) demonstrate that Runge-Kutta methods have poor stability properties. In addition, the semi-explicit property whereby the differential and algebraic equations cannot be combined and are non-linearly dependant on one-

another creates a situation where the internal Runge-Kutta solutions within the timestep Δt must also be coupled with algebraic solutions at the same stage.

We begin this work by introducing the following numerical (Euler) discretization of Equation (1) as:

$$\frac{X^{n+1} - X^n}{\Delta t} = \theta f(X^{n+1}, Y^{n+1}) + (1 - \theta)f(X^n, Y^n) + \mathcal{O}(\Delta t^k) \tag{2}$$
$$0 = g(X^{n+1}, Y^{n+1}, t)$$

where; $\theta = 0.0$ yields a fully-explicit temporal discretization, yielding a discretization error of order $k = 1$ (i.e. first-order accurate. Note that "fully-explicit" now refers to the actual Euler method used to discretize the time derivative, and not the form of the DAE. Superscripts $n + 1$ and n denote the current and past time solutions separated over an increment of time Δt.

PHYSICAL INFRASTRUCTURE SECTOR

Formulation and parameterization of the physical infrastructure sector follows from Rehan et al. (2013a,b) and is not repeated here for brevity. This section presents that component of the system dynamics model shown in Figure 7 of Rehan et al. (2013a) belonging to the physical infrastructure sector, with the presentation involving primary and secondary variables following Equation (2) above. Note that this system dynamics model also includes finance and consumer sectors which are omitted here for brevity. In fact, secondary variables within the physical infrastructure sector that provide links to the finance and consumer sectors are also omitted for brevity.

Primary variables and equations. Degradation of pipes is captured by the transient accumulation and subsequent depletion of pipe lengths in five different age groups over the course of the simulation period. This process is captured by five primary variables denoted as: $SewersCondition_{M, ICG_i}^{n+1} [km]$ where M denotes material (vitrified clay, concrete, PVC) and ICG_i denotes condition grade $i \in \{1 \dots 5\}$ where "1" is new installation and "5" is completely deteriorated. These $i \times M$ primary variables lead to differential equations of the following form:

$$\frac{SewersCondition_{M, ICG_i}^{n+1} - SewersCondition_{M,, ICG_i}^{n}}{\Delta t}$$
$$= \theta\left(Deterioration_{M, ICG_{i-1,i}}^{n+1} - Deterioration_{M, ICG_{i,i+1}}^{n+1}\right)$$
$$+ (1 - \theta)\left(Deterioration_{M, ICG_{i-1,i}}^{n}\right) \tag{3}$$
$$- Deterioration_{M, ICG_{i,i+1}}^{n}\right)$$

Secondary variables and equations. The ageing process of the pipes is facilitated thought the set of secondary variables denoted as $Deterioration_{M, ICG_{i,i+1}}^{n+1}[km/yr]$. These secondary variables are quantified as:

$$Deterioration_{M, ICG_{i,i+1}}^{n+1} = \frac{SewersCondition_{M, ICG_i}^{n+1}}{AvgDurationInGrade_{M, ICG_i}} \tag{4}$$

The average time that a pipe of a specific material spends in each ICG group is defined as $AvgDurationInGrade_{M,\ ICG_i}\ [yr]$ and is a constant. These links take the form of algebraic equations and specify non-linear feedbacks and controls on the temporal evolution of these sectors.

RESULTS

In the context of this work, we restrict our attention to discussing the significance of any potential numerical errors in the solution of the Rehan et al. (2013a, 2013b) system dynamics problem. Note that interpretation of the physical relevance of the simulation results to the financially sustainable management of a watewater distribution network was the subject of Rehan et al. (2013a, 2013b) and is not repeated here for brevity. Initially, we would like to ensure that the DAE algorithm is point-wise stable in that they converge to a single solution as the timestep is progressively reduced. To illustrate this discussion, Figures 2 and 3 present some primary and secondary variables over the 100-year simulation period with a logarithmic time scale to focus on the key dynamics. The largest timestep is $\Delta t = 2^{-2}\ year$ to conform to the Rehan et al (2013a, 2013b) simulation results. The smallest timestep STELLA 7.0.2 would accept was $\Delta t = 2^{-8}\ year$. For the numerical DAE approach, the timestep was progressively halved to $\Delta t = 2^{-9}\ year$.

Figure 1a shows the length of the pipes in each condition group over the simulation period. Clearly, both Runge-Kutta and the numerical DAE approach produced visually-identical answers over the entire simulation period. This result appears to be largely a consequence of the fact that the infrastructure sector is the sector with having minimum nonlinearity in the equations. We conclude that the Runge-Kutta approach employed by STELLA 7.0.2 is virtually identical to a numerical DAE approach for linear systems.

Figure 1b presents simulation results for the primary variable $FundBalance^{n+1}$. These results indicate that the funds balance does deviate from "zero" despite the intent of Rehan's model to signal financial sustainability though a zero funds balance.

Figures 1c and 1d present simulation results for the primary variables $SewageFee^{n+1}$ and $WaterDemand^{n+1}$, respectively. It is seen that at early times there is oscillation in the variable $SewageFee^{n+1}$, but as time goes on the graph becomes smoother. Not very much difference is noticeable in STELLA's results by changing the timestep size unlike the DAE scheme. As the timestep is refined, neither Runge-Kutta nor the numerical DAE solutions appear to converge to the same solution. However, the numerical DAE solution does exhibit the desired point-wise stability behavior of converging to a single solution over the entire simulation period as the timestep is refined.

Figure 2a and 2b depict $SewageFeeHike^{n+1}$ and $SewageFeeDecline^{n+1}$ which is set based on the Required Sewage Fee to cover all the operational expenditures and part of the capital expenditures of that year. Figures 2c and 2d depict $Revenue^{n+1}$ and $Debt^{n+1}$ respectively. $Revenue^{n+1}$ is the product of $SewageFee^{n+1}$ and $WaterConsumption^{n+1}$.
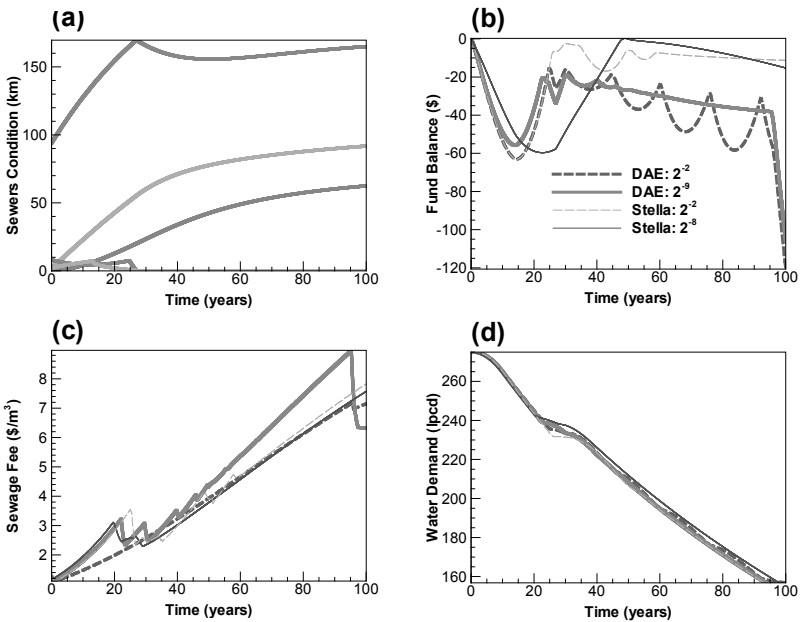
Figure 1. Primary variables over the 100-year simulation period. All numerical DAE results obtained using Fully Explicit method. The timestep size Δt for each line is enumerated in the legend, with DAE simulation results provided by thick lines and numerical STELLA results by thin lines.

As the timestep is refined the results change significantly from the results obtained by picking a course timestep size. $Debt^{n+1}$ is the amount of money that the utility can borrow and is a function of the amount of deficit of the network in covering its Capital Expenses. Furthermore, the system is allowed to borrow money if its capacity for borrowing has not exceeded a pre-defined value. Figures 2e and 2f depict the $OpEx^{n+1}$ and $CapEx^{n+1}$ (operational and capital expenditures) respectively.

To demonstrate that the numerical DAE solution is point-wise stable and converging to the correct solution, we analyze the convergence for the $SewageFee^{n+1}$ and $FundBalance^{n+1}$ primary variables at year 80 within the 100-year simulation period in Figure 3. This point in time was chosen because the Runge-Kutta and the numerical DAE solutions show a clear divergence in their values.

In summary, the DAE approach is point-wise stable, since as the timestep size is refined the numerical solution is converging to a single answer. STELLA's results are seen to be insensitive to the timestep size.
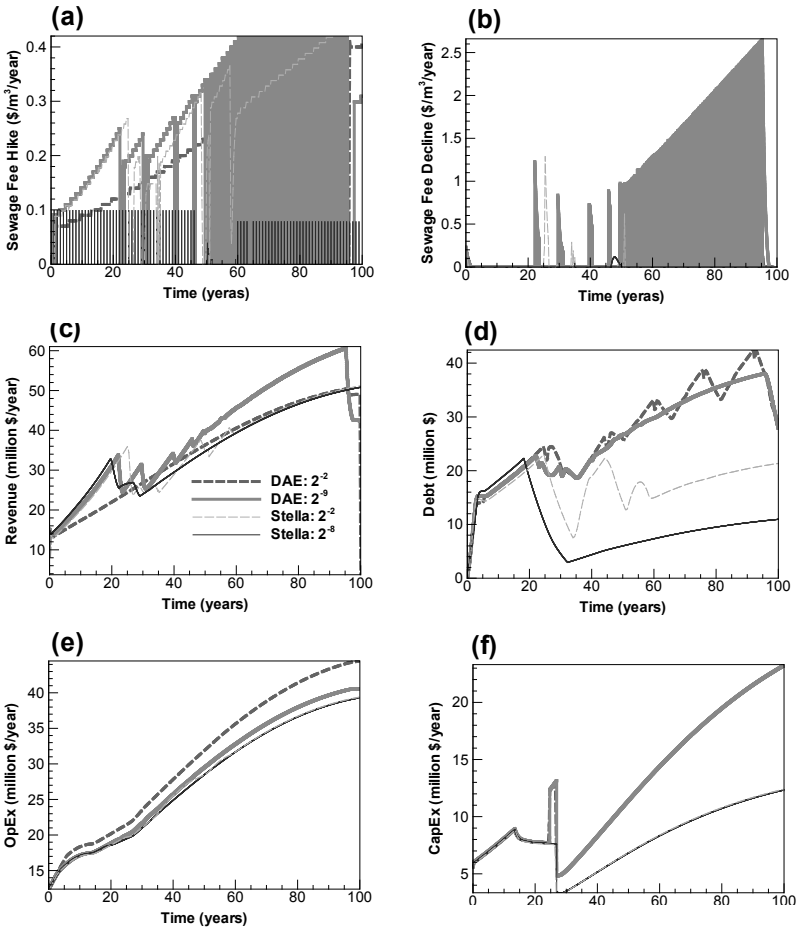
Figure 2. Secondary variables over the 100-year simulation period. All numerical DAE results obtained using Explicit Euler method. The timestep size Δ*t* for each line is enumerated in the legend, with DAE simulation results provided by thick lines and numerical STELLA results by thin lines.
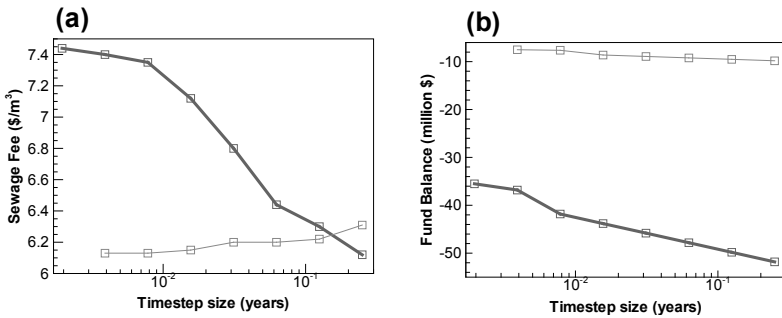
Figure 3. Convergence of the numerical DAE and STELLA 7.0.2 algorithms as the timestep is sequentially halved, for the primary variables $SewageFee^{n+1}$ and $FundBalance^{n+1}$ at year 80.

CONCLUSIONS

The system dynamics model of Rehan et al. (2013a), which focuses on modeling of a financially sustainable wastewater distribution network using the software package STELLA, provides the prototype problem for this work. Specifically, we illustrate a strategy for converting system dynamics mathematical vernacular into a semi-explicit index-1 DAE system. The differential equations capture the transient evolution of the modeled system via a series of independent primary variables that all include a time derivative. In addition, a series of algebraic (secondary) equations and variables follow that are used to specify the non-linear behavior of the system as well as any controls on the physically relevant range (and hence temporal evolution) of the primary variables. In the case of the Rehan et al. (2013a) system dynamics problem, a secondary variable is identified that specifies the non-linear behavior by which consumers adjust their change in demand for water as its unit price increases, with this demand change involving the price elasticity of demand for water. Also, secondary variables are identified that specify the rate at which the utility should adjust the unit price of water in order to achieve financial sustainability, expressed by maintaining a zero funds balance with revenues equalling expenditures. These variables act as a control on the evolution of the system ensuring that the fee hike (and decline) rate on the unit cost of water never exceeds an upper bound presumably set by the municipal council, while maintaining the funds balance within specified upper (surplus) and lower (deficit) bounds. The set of equations is solved by discretizing the time derivatives using fully-explicit temporal weighting in a numerical Euler-based DAE scheme with a fixed-point iteration to resolve the non-linearity.

Values of primary and secondary variables are presented over the 100-year simulation period of the Rehan et al. (2013b) system dynamics problem using the numerical DAE approach as well as the Runge-Kutta algorithm invoked by the STELLA software package. Both algorithms deviate markedly once the non-linearity of the system becomes dominant and the control is required to enforce a zero funds

balance. Analysis of the numerical DAE solution at this point of departure indicate that it is point-wise stable in the sense that the primary variables converge to a single solution as the timestep is progressively reduced. We therefore conclude that system dynamics software packages are invaluable tools to help the civil engineer develop prototype numerical models of complex systems. However, it is the due diligence of the civil engineer to employ accurate numerical algorithms to ensure that solutions to these complex systems are free from numerical aberrations.

REFERENCES

Ascher, U.M., Petzold, L.R., 1998. Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations. SIAM.

Bendtsen, C., Thomsen, P.G., 1999. Numerical Solution of Differential Algebraic Equations. Technical Report, Department of Mathematical Modelling, Technical University of Denmark. http://www2.imm.dtu.dk/documents/ftp/tr99/tr08_99.pdf accessed 14.01.13.

Brenan, K.E., Campbell, S.L., Petzold, L.R., 1989. Numerical Solution of Initial-Value Problems in Differential Algebraic Equations, New York: North-Holland, p266.

Forrester, J.W., 1958. Industrial dynamics: a major breakthrough for decision makers. Harvard Business Review 36 (4), 37-66.

Petzold, L.R., 1982. Differential /Algebraic Equations Are Not ODE'S, 1982. SIAM Journal on Scientific and Statistical Computing 3 (3), 367-384.

Rehan, R., Knight, M.A., Haas, C.T., Unger, A.J.A., 2011. Application of system dynamics for developing financially self-sustaining management policies for water and wastewater systems, Water Research, 45 (16), 4737-4750.

Rehan, R., Knight, M.A., Haas, C.T., Unger, A.J.A., 2013a. Financially sustainable management strategies for urban wastewater collection infrastructure – Development of a system dynamics model, Tunnelling and Underground Space Technology, Article in Press, http://dx.doi.org/10.1016/j.tust.2012.12.003

Rehan, R., Knight, M.A., Haas, C.T., Unger, A.J.A., 2013b. Financially sustainable management strategies for urban wastewater collection infrastructure – Implementation of a system dynamics model, Tunnelling and Underground Space Technology, Article in Press, http://dx.doi.org/10.1016/j.tust.2012.12.003

Seppelt, R., Richter, O., 2005. "It was an artefact not the result": A note on system dynamics dynamic model development tools. Journal of Environmental Modelling & Software, 20 (12), 1543-1548.

Son, J., Rojas, E.M., 2011. Impact of optimism bias regarding organizational dynamics on project planning and control. Journal of Construction Engineering and Management, ASCE, 137 (2), 147-157.

Rizzo, D.M., Mouser, P.J., Whitney, D.H., Mark, C.D., Magarey, R.D., Voinov, A. A., 2006. The comparison of four dynamic systems-based software packages: Translation and sensitivity analysis. Journal of Environmental Modelling & Software, 21 (10), 1491-1502.

Using a Life Cycle Assessment Approach for Optimizing Multi-Objectives in Construction Projects

Gulbin Ozcan-Deniz[1] and Yimin Zhu[2]

[1] Assistant Professor, Department of Civil and Architectural Engineering and Construction Management, Milwaukee School of Engineering, Milwaukee, WI; 53202; PH (414) 277-7205; FAX (414) 277-7415; Email: ozcan@msoe.edu
[2] Associate Professor, Department of Construction Management, Florida International University, Miami, FL, 33174; PH (305) 348-3517; FAX (305) 348-6255; Email: zhuy@fiu.edu

ABSTRACT

With environmental concerns as another dimension of constraints in the built environment, life cycle assessment (LCA) has gained tremendous importance. It has been commonly accepted among researchers that environmental impacts of construction operations need to be analyzed and monitored throughout the whole life cycle, including raw material extraction, manufacturing, transportation, construction, and end-of-life steps. Furthermore, life cycle approach has brought the concept of systems thinking into the built environment to understand the interactions among various systems and elements of these systems. Previous studies have focused on modeling construction projects by using systems thinking. This research builds on these studies in using the results of construction system models to optimize multiple project objectives. Environmental impact is set as a new project objective to reflect environmental concerns throughout the project life. Time and cost are selected to be optimized with environmental impact considering their importance for owners and contractors. Multi-objective optimization is used as an effective tool to optimize these three potentially conflicting objectives in construction projects. This paper aims to model the optimization process of time, cost and environmental impact (TCEI) with the help of an LCA approach and multi-objective optimization techniques. LCA will allow observing and analyzing project objectives (TCEI) throughout the life cycle of the project, while the multi-objective optimization procedure will help to view the interdependency of time, cost and environmental impact by considering the life cycle of construction projects.

INTRODUCTION

The construction sector in the U.S is accounted for 39% of primary energy use and 38% of CO_2 emissions (USGBC, 2008). The considerable share of construction industry on the adverse environment impacts has raised concerns in the built environment. Consequently, sustainability and green building concepts have been

formed to minimize the harmful consequences of construction activities. According to USGBC (2012), the three major concepts of sustainability are systems thinking, life cycle thinking, and integrated process. Life cycle assessment (LCA), which is a tool to calculate environmental impact of products starting from extraction to disposal, has been used for quite a while to calculate the environmental impacts of construction activities. Additionally, system dynamics (SD), a methodology for studying and managing complex systems (Sterman, 2000), has been utilized to understand the interactions among various systems and elements of these systems. The environmental impact of buildings and their operations have received a significant amount of research attention, e.g., optimization of building operations (Zhu, 2006). With the help of LCA and SD, environmental impact can be modeled to analyze and optimize various project objectives. Environmental impact being one of the project objectives allows the proposed model to reflect environmental concerns throughout the project life. In this study, time, cost and environmental impact (TCEI) will be analyzed to view the interdependency between them. Being an effective tool in optimizing multiple parameters, multi-objective optimization is used in this study to optimize these three potentially conflicting objectives in construction projects. The first part of this paper will give the literature information about LCA and multi-objective optimization. Second part will focus on the details of the proposed model. Then, an example case study will be analyzed and results will be evaluated. Finally, conclusions and future directions regarding this study will be summarized.

LITERATURE REVIEW

LCA is an important tool for evaluating the environmental impacts of products and services from a "cradle to grave" perspective (Bengtsson, 2001). It is described as the most sophisticated tool to consider and quantify the consumption of resources and the environmental impacts associated with a product or process (Curran, 1996). Life cycle includes material exploitation, material production, construction & installation, operation & maintenance, and demolition phases (Zhang, Hammad and Bahnassi, 2009).

Environmental impacts can be calculated in different categories such as green house gas (GHG) emissions, energy consumption, acidification, pollutants to air and water, etc. through the life cycle phases. Construction literature is rich in LCA studies that focus on energy consumptions and GHG emissions of buildings. For example, Sartori and Hestnes (2007) used LCA to compare the energy utilization of conventional and low-energy buildings through their lives. In a similar manner, Xing et al. (2008) applied a life-cycle inventory model to compare two construction materials: steel and concrete, in terms of energy consumptions and emissions. LCA has been widely applied to the evaluation of building products, systems, and construction processes, and considered as a sophisticated tool to quantify the consumption of resources and environmental impact associated with a product or process (Curran, 1996). In the scope of this research, the calculation of environmental impacts is also focused on evaluating GHG emissions by using the LCA approach.

Although the environmental optimization is a developing concept in construction, it is not easy to adopt environmental management in construction since it does not cause a reduction in the total time or cost of the project and inversely the environmental regulations can give rise to delay or cost overrun. In a similar manner, any limitation in time or cost can result in an increase or decrease in the environmental impacts of the project (Chen, Li and Wong, 2000). There has not been such an optimization tool to handle the interrelations between these three objectives in construction projects. As a recent sustainable optimization example Wu et al. (2010) performed multi-objective optimization analysis on the design of water distribution systems considering economic (life cycle cost) and environmental objectives. They also used genetic algorithms and concluded that significant trade-offs between the economic and environmental objectives. Another example is Marzouk et al. (2008) tried to optimize total pollution with project duration and cost by using genetic algorithms. They analyzed the dynamic nature of construction activities referring to different types of relationships and the change of activities' criticality and proposed a framework to minimize time, cost and pollutants of construction projects. Considering the previous optimization studies, there is still a need for not only integrating TCEI of construction operations but also handling EI from a life cycle perspective. The multi-objective optimization theory in literature is enlarged in this paper by adding environmental impact (EI) as a new performance parameter to the traditional project criteria, which are time (T) and cost (C).

METHODOLOGY

A multi-disciplinary approach is used to create the framework for multi-objective optimization of TCEI. The flow chart for the proposed framework is shown in Figure 1. As mentioned before, this study is based on the results of the previous SD simulation (Ozcan-Deniz and Zhu, 2012). The previous SD model was basically used to determine a construction method that can best satisfy three potentially conflicting objectives as TCEI. Constraints that are imposed by external factors, like project conditions or internal factors (e.g., project participant relationships) are modeled in the construction method selection. The probability of occurrence of project conditions are entered as inputs and the most feasible construction method was output based on the combination of material selection, equipment selection, and stabilizing additive selection values. Previous SD model has supplied various construction project models that can create simulation scenarios regarding different probability inputs. Then, TCEI are calculated based on these different scenarios.
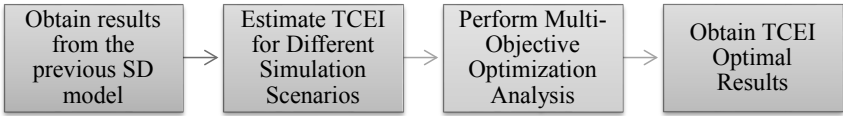
Figure 1. Flow Chart for the Proposed Framework.

Three types of data such as project duration, project cost and environmental impact were collected and processed before they are used for optimization and analysis. Data in the form of plans, contract (which includes resource types and amount as well as fuel consumptions of equipments), schedule, and daily progress report were collected from the Florida Department of Defense (FDOT). Then, the required information was identified through these sources and time, cost and environmental impact values were calculated for alternate construction operations (including the original construction method). Project duration and cost data was derived using conventional scheduling and estimating methods. Environmental impacts were limited to GHG emissions, which were calculated based on data from literature. Environmental impact data of observed or alternate construction operations could only be derived by using life cycle assessment (LCA) due to the lack of data support in practice.

The multi-objective optimization procedure is designed based on literature and the outcomes of the proposed SD model to look for the relationship between TCEI. EI is limited to greenhouse gas emissions (GHG) in this study. The multi-objective optimization process searches for optimal/near optimal trade-offs among minimizing project duration, project cost, and project GHG. Therefore, the objective function is expressed as in Equation 1.

Objective function = Minimize (Project duration + Project cost   **Equation 1**
+ Project GHG)

To understand the multivariable approach, a Euclidian space is defined having the variables T, C and EI. Each solution of project objectives can be plotted in the three-dimensional space, the origin of which has (0, 0, 0) values for TCEI. The solutions are expected to lie in a certain boundary called the possibility frontier. The possibility frontier in the space created with the lowest boundary being the corner of the box created by the boundaries that is closer to the origin. This point is calculated by selecting the minimum time, cost and environmental impact achievable in each activity and adding them accordingly to obtain the total minimum achievable. The new space of possible solutions is normalized using as an upper limit of the corresponding maximums of time, cost and environmental impact. The normalized values for time, cost and environmental impact for a solution K are calculated by using Equation 2, Equation 3, and Equation 4. The objective function utilizes the normalized values of the project objectives to find the optimal TCEI values. The fitness function in Equation 5 is used to evaluate the performance of each solution.

The performance is determined by a fitness function, which is defined as the total distance between the solution obtained and the origin in the new normalized subspace. A lower fitness value means a better solution as the solution is closer to the origin.

$$Normal\ T_K = \frac{T_K - T_{min}}{T_{max} - T_{min}}$$ **Equation 2**

$$Normal\ C_K = \frac{C_K - C_{min}}{C_{max} - C_{min}}$$ **Equation 3**

$$Normal\ EI_K = \frac{EI_K - EI_{min}}{EI_{max} - EI_{min}}$$ **Equation 4**

$$f\ (T_K, C_K, EI_K\ ) = \sqrt{Normal\ T_K{}^2 + Normal\ C_K{}^2 + Normal\ EI_K{}^2}$$ **Equation 5**

RESULTS AND ANALYSIS

A population of solutions is generated randomly by using the results of the previous SD model (Ozcan-Deniz and Zhu, 2012). The values shown in Table 1 are used as a starting point in optimization.

Table 1. TCEI Results

Trial #	T	C	GHG
1	36	$1,945,751.89	3,857,983.11
2	36	$1,945,751.89	3,857,983.11
3	36	$1,998,046.87	4,130,521.55
4	36	$1,830,894.80	2,704,097.84
5	36	$1,830,894.80	2,704,097.84
6	36	$1,832,741.60	2,704,097.84
7	36	$1,961,275.87	4,556,064.96
8	36	$1,832,741.60	2,704,097.84
9	36	$1,832,741.60	2,704,097.84
10	36	$1,945,751.89	3,857,983.11
11	22	$1,363,662.60	3,313,032.61
12	22	$1,363,662.60	3,313,032.61
13	36	$1,961,275.87	4,556,064.96
14	36	$2,026,281.78	4,466,841.63
15	36	$2,026,281.78	4,466,841.63
16	36	$1,832,741.60	2,704,097.84
17	36	$1,945,751.89	3,857,983.11
18	36	$1,961,275.87	4,556,064.96
19	36	$2,026,281.78	4,466,841.63
20	22	$1,363,662.60	3,313,032.61

The normalized TCEI values are calculated based on the random generated population of solutions. Then, the one or ones having the minimum objective function,

i.e. being closer to the origin of a new subspace, is selected to be in the optimal solution set. Simple linear correlation, also referred to as Pearson correlation, is used to determine the extent of relationship between the three project objectives (TCEI). Pearson correlation compares two variables at once, and the value of correlation gives if the variables are proportional. If they are proportional, that means they are linearly related. The value of correlation or correlation coefficient gets values between -1.00 and 1.00, where they represent negative and positive correlation respectively. Pearson's correlation coefficient is calculated by using three different sums of squares (SS). As there are three variables (time, cost, and EI) in this study, there are three sets of SS calculation in pairs (time & cost, time & EI, and cost & EI). The SS formulas for variables and the cross product are shown in Equation 2, Equation 3, and Equation 4 for time & cost. These equations are further used to find the correlation coefficient (r), which is shown in Equation 5. By following the same steps and using similar equations, the correlation coefficient is calculated for time & EI and cost & EI. The SS values and correlation coefficients of TCEI are shown in Table 2 and Table 3.

$$SS_{TT} = \sum_{i=1}^{i=n}(T_i - \bar{T})^2$$

Equation 6

$$SS_{CC} = \sum_{i=1}^{i=n}(C_i - \bar{C})^2$$

Equation 7

$$SS_{TC} = \sum_{i=1}^{i=n}(T_i - \bar{T})(C_i - \bar{C})$$

Equation 8

$$r = \frac{SS_{TC}}{\sqrt{(SS_{TT})(SS_{CC})}}$$

Equation 9

Table 2. SS Values of TCEI Pairs.

Time & Cost	Values	Cost & EI	Values	Time & EI	Values
SS_{CC}	8.99E+11	SS_{CC}	8.99E+11	SS_{TT}	499.80
SS_{TT}	499.80	SS_{EIEI}	1.06E+13	SS_{EIEI}	1.06E+13
SS_{TC}	2.01E+07	SS_{CEI}	1.48E+12	SS_{EIEI}	1.37E+07

Table 3. Correlation Coefficients of TCEI.

r	C	T	GHG
C	1		
T	0.9466	1	
GHG	0.4789	0.1887	1

The value of the correlation coefficient defines two properties of the correlation: (1) The sign of r gives if the correlation is negative or positive; (2) The magnitude of r gives the strength of correlation. Between time & cost, r = 0.9466 suggests a strong positive correlation. When the correlation coefficient is between 0.3 and 0.7 as in the

case of cost & EI, it is interpreted as a moderate positive correlation. Between time & EI, the correlation coefficient is below 0.3, which suggests a weak positive correlation.

While the correlation coefficient (r) represents the linear relationship between two variables, the coefficient of determination (r^2) represents the proportion of common variation in the two variables. For example, the coefficient of determination of 0.90 between time & cost suggests that 90% of the variability in the cost of the project is explained by its duration. Similarly, the coefficient of determination of 0.23 between GHG & cost suggests that 23% of the variability in the cost of the project is explained by its GHG, and the coefficient of determination of 0.04 between time & GHG suggests that only 4% of the variability in the GHG of the project is explained by its duration.

CONCLUSIONS

The multi-objective optimization procedure in this study resulted in remarkable outputs regarding the interdependency between TCEI. First, the coefficient of determination of 0.90 between time & cost suggests that there is a decrease in the project cost, when there is usually a decrease in the project duration. This is something that can be observed from the data set. The reason for 90% being different from generally accepted time-cost tradeoff lies beneath the construction methods. The new technology in delivering various construction activities causes a decrease in both time & cost, so that they have a common variation. Additionally, switching from fuel types or equipment types not only affects GHG, but also shows the same kind of behavior for time & cost. This situation of having the same increase or decrease, time & cost shows consistency in their pattern.

Secondly, the coefficient of determination of 0.23 between GHG & cost suggests that 23% of the variability in the cost of the project is explained by its GHG. The common variation between GHG & cost is low, as the GHG data does not always increase when cost increases. The common variation between time & GHG is even lower.

Finally, the coefficient of determination of 0.04 between time & GHG suggests that only 4% of the variability in the GHG of the project is explained by its duration. The very low variability can be explained by the data set, as the data of time & GHG does not show the same pattern of behavior. The duration of project based on different construction methods can have the same value, while having the same GHG value is only limited to certain construction methods.

To sum up, the correlation coefficient (r) and the coefficient of determination (r^2) help contractors to visualize the interdependency of time, cost and environmental impact by considering the life cycle of construction projects. However, as GHG & cost and time & GHG do not have significant correlations, the relationships between them are case-by-case. For each case, the variability between each pair can change. Therefore,

whether or not there is a certain relationship between these objectives cannot be generically set.

ACKNOWLEDGEMENTS

REFERENCES

Bengtsson, M. (2001). "Weighting in practice: implications for the use of life-cycle assessment in decision making", *Journal of Industrial Ecology*, 4(4), 47-60.

Chen, Z., Li, H., and Wong, C. T. C. (2000). "Environmental management of urban construction projects in China", *Journal of Construction Engineering and Management*, 126, 320-324.

Curran, M. A. (1996) *Environmental Life-Cycle Assessment*, NY: McGraw-Hill.

Curran, M. A. (1996) *Environmental Life-Cycle Assessment*, NY: McGraw-Hill.

Marzouk, M., Madany, M., Abou-Zied, A., and El-Said, M. (2008). "Handling construction pollutions using multi-objective optimization", *Construction Management and Economics*, 26, 1113-1125.

Ozcan-Deniz, G., and Zhu, Y. (2012). "Evaluating Construction Methods for Low Carbon Emissions Using System Dynamics Modeling", *International Conference on Computing in Civil Engineering*, Clearwater Beach, FL, 610-617.

Sartori, I., and Hestnes, A. G. (2007). "Energy use in the life cycle of conventional and low-energy buildings: A review article", *Energy and Buildings*, 39, 249-257.

Sterman, J. (2000) *Business dynamics: Systems thinking and modeling for a complex world*, New York: Irwin McGraw-Hill.

USGBC (2008) *A National Green Building Research Agenda*, [Online], Available: http://www.usgbc.org/ShowFile.aspx?DocumentID=3402 [16 December 2010].

USGBC (2012) *Green Building and LEED Core Concepts Guide*, 2nd edition, Washington, DC: U.S. Green Building Council.

Wu, W., Simpson, A. R., and Maier, H. R. (2010). "Accounting for Greenhouse Gas Emissions in Multiobjective Genetic Algorithm Optimization of Water Distribution Systems", *Journal of Water Resource Planning and Management*, 136(2), 146-155.

Xing, S., Xu, Z., and Jun, G. (2008). "Inventory analysis of LCA on steel and concrete construction office buildings", *Energy and Buildings*, 40, 1188-1193.

Zhang, C., Hammad, A., and Bahnassi, H. (2009). "Collaborative multi-agent systems for construction equipment based on real-time field data capturing", *Journal of Information Technology in Construction (ITcon), Special Issue Next*

Generation Construction IT: Technology Foresight, Future Studies, Roadmapping, and Scenario Planning, 14, 204-228.

Zhu, Y. (2006). "Applying computer-based simulation to energy auditing: a case study", *Energy and Buildings*, 38, 421-428.

An Expert System Based on OpenStudio Platform for Evaluation of Daylighting System Design

Jia Hu[1] and Svetlana Olbina[2]

[1]PhD Candidate, Rinker School of Building Construction, University of Florida, Gainesville, FL 32611-5703; PH (352) 273-1175; email: hujia@ufl.edu
[2]Assistant professor, Rinker School of Building Construction, University of Florida, Gainesville, FL 32611-5703; PH (352) 273-1166; email: solbina@ufl.edu

ABSTRACT

Simulation tools can provide valuable information for daylighting system design while optimization methods can determine feasible solutions from a number of uncertain design parameters. However, optimization methods allow for limited interactivity and only consider several performance criteria. Therefore, to utilize experts' domain knowledge, this research presents a rule-based expert system based on OpenStudio platform, which integrates daylighting and thermal simulations. The expert system can interpret and compare different design changes and options based on the pre-defined rules and simulation results, and can provide possible solutions in the case of failure to meet design performance criteria. The users can interactively provide alternative design options to the expert system for evaluation based on users' knowledge. The expert system can provide guidance to evaluate daylighting design.

INTRODUCTION

The design metrics (e.g., daylighting and thermal metrics) need to be measured during the design of a daylighting system to evaluate the effectiveness of the design parameters for the specific daylighting system. Parameters considered in the design of a daylighting system may include, for example, window dimensions and position, glazing transmittance, and overhang dimensions. Various computer-aided daylighting design tools were developed to optimize these design parameters (Crawley et al. 2001; Hu and Olbina 2012; Reinhart and Wienold 2011). However, some problems still exist. For example, these tools did not effectively utilize designers' knowledge although designers have an important role in the design of daylighting systems. Optimization techniques were integrated into some design tools to optimize design parameters. However, designers cannot simply accept the design calculated by these optimization programs because the optimization programs rely on only a few design metrics. Another problem is that the parametric design method may require a long time to calculate design metrics.

Instead of solely relying on the parametric design scheme, a better approach may be the one that combines optimization methods with designers' domain knowledge in order to better utilize designers' knowledge. Thus, some expert systems were developed to utilize experts' knowledge about a specific domain that was

encoded in an algorithm or a compute system (Guo et al. 1993; Jung et al. 2003). The expert system uses inference rules usually expressed as If-Then statements and fires chain of rules to calculate results (Gonzalez and Dankel 1993). Some of the rules were used for selection of lighting system (Jung et al. 2003) or adjusting the design goals (Gagne et al. 2011; Jung et al. 2003). Recently Gagne et al. (2011), developed an interactive expert system using Lightsolve Viewer (LSV) simulation engine for lighting simulations. LSV formulated a part of the knowledge base of the expert system. A designer can interactively adjust some facade parameters (e.g., window position, window dimension) by using fuzzy rules of the expert system. The problem was that if a designer changes design parameters lighting and energy simulations had to be conducted again to provide simulation results. In addition, existing expert systems did not provide information about the impact of design parameters on the daylighting and thermal performance. The problem with the rule expressions in some expert systems is that they do not provide sufficient information for adjusting daylighting parameters.

Therefore, the aim of this research was to develop a framework of the rule based expert system based on simulation data. The expert system could help designers to design window dimensions and location, glazing types, overhang dimensions and location. The research objectives were to: (1) develop a user interface that could interacts with designers; (2) develop set of rules for the expert system; and (3) develop a workflow of daylighting and energy simulations from which the knowledge can be derived and stored in the database.

MODEL DEVELOPMENT

The overall structure of the expert system shown in Figure 1 includes two main modules: the user interface and the expert system engine. The user interface was implemented in the SketchUp/OpenStudio platform, and used for user input and for demonstrating results returned from the expert system. The returned results would be 3D distributions of the illuminance, daylight autonomy (DA) (Nabil and Mardaljevic 2006), and window heat gain/loss. DA denotes a percentage of occupied times of a year when the minimum required illuminance level (500 lx in this research) at a point in a space is provided by daylight only. The figures can be generated by Python Matplotlib (Hunter et al. 2012).

The module of the expert system engine includes two submodules: knowledge database and rule inference engine. The knowledge database stores information of simulation results generated by daylighting and energy simulations as well as user input. The rule inference engine includes three rule sets that are used to create and modify information stored in the knowledge database.

The module of the expert system engine was mainly implemented in CLIPS and Python. CLIPS was used for the rule execution and Python was used to control simulations and generate input files. EnergyPlus and Radiance can be used for energy and daylighting simulations to create the knowledge database. The simulation results would be translated to the knowledge representation format recognized by CLIPS.

Three rule sets were developed (see Figure 1). The rule set 1 is to select an appropriate base building model with simulated data from the knowledge database. These data will be used by the rule sets 2 and 3 to evaluate the impact of the change

of design parameters on the daylighting (e.g., DA) and thermal (e.g., window heat gain/loss) performance. Then 3D distributions of daylighting levels and 3D distributions of window heat gain/loss can be generated by comparing the user-defined model with the base building model selected by the rule set 1. The data are stored in the knowledge database and can be retrieved by the rule engines during the execution. The data needed by the three rule sets include the daylighting and thermal performance of base building models and building parameters. These data should be generated by EnergyPlus and Radiance beforehand so that the user can reduce the simulation time.
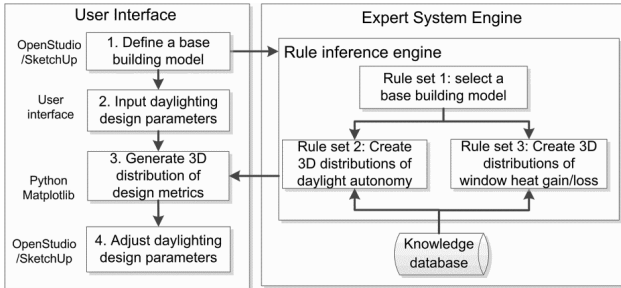
Figure 1. The overview of the structure of the expert system

User Interface

The user interface module includes four steps (see Figure 1). In Step 1, the user needs to create a building model using SketchUp plugin and OpenStudio. The parameters of the building model include the building form, and thermal and daylighting properties of materials. The thermal properties should be in accordance with the ASHRAE standards. The typical building parameters have been integrated into OpenStudio. Some other parameters such as dimensions and location of shading systems have to be determined by designers. The aim of the expert system is to help designers to optimize design parameters of these shading systems.

In Step 2, designers should input daylighting design parameters through the user interface developed in SketchUp by using OpenStudio (see Figure 2). The input data provided by users will be used by the rule inference engine and knowledge database. The user interface includes six items:

    (1) Sensor density (distance between sensor points);
    (2) Design metric (three possible options are: daylighting, energy, and daylighting and energy);
    (3) Sky type (four possible options of sky models are: CIE overcast, CIE clear, CIE intermediate, and typical actual sky from weather file);
    (4) Time period (the time periods that can be used for daylighting simulations are: a particular time of a day, one day, or whole year);
    (5) Design parameter (daylighting design parameter that should be optimized are: glazing transmittance; window dimensions; window location; overhang width and overhand depth);
    (6) Parameter change (the maximum percentage of a change of the design parameter value in Item (5)).

In Step 3, the expert system can provide 3D distribution figures showing how the different percentage changes of a design parameter (i.e., Menu item (5) in Figure 2) influence the design metric (i.e., Menu item (2) in Figure 2). An example of such a 3D figure is shown in Figure 3, which was generated by Python Matplotlib allowing users to zoom and rotate the figure. The X axis represents the space width, the Y axis represents the space depth, and the Z axis represents the values of daylight autonomy at different sensor points.
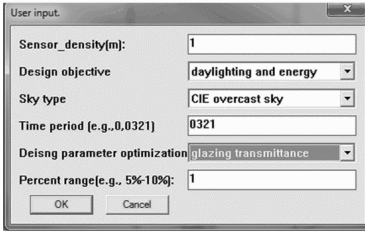
Figure 2. The user input of the expert system
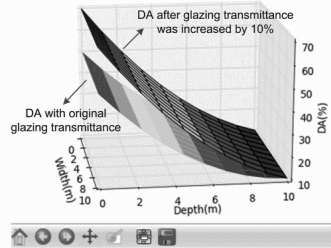
Figure 3. An example of 3D distributions of daylight autonomies (DAs) for two different values of glazing transmittance

In Step 4, designers can make final decisions using both their own knowledge and the information provided by the expert system, and adjusts the design parameter specified in Menu item (5) in Figure 2 in SketchUp. For example, a designer may decide to change window glazing transmittance if the overall 3D distribution figure shows that the values of DA are too low.

Expert system engine: Knowledge database module

The knowledge database stores user input and pre-simulated data of energy (i.e., window heat gain/loss) and daylighting levels. The format of the pre-simulated data can be represented by Eq. (1).

$$D = \{sky, climate, ref, geom, thermal, shading, energy, daylighting\} \quad (1)$$

Where D = data representation stored in the knowledge database; sky = sky model (CIE clear, overcast, intermediate sky models, or actual sky); $climate$ = ASHRAE climate zone; ref = building interior reflectance; $geom$ = building space geometry; $thermal$ = thermal properties based on ASHRAE baseline buildings; $shading$ = design parameters of shading systems (e.g., window dimensions and position); $energy$ = window heat gain/loss; $daylighting$ = daylighting illuminance at sensor points.

The whole process of generating these simulation data and data representation in Eq. (1) includes four steps (see Figure 4). In Step 1, most of the building model parameters will be defined (see Figure 4). The parameters in Eq. (1) are discrete and thus different combinations of these discrete values could be selected. In Step 2, a number of sub-parameters of the shading will be setup. This would be introduced in the later part of this section. In Step 3, a number of different values of window-to-wall ratios (WWRs) will be selected: 30%, 50%, 60%, 70%, 80% and 90%. The energy data are less affected by the window location. For each building with the same WWR, two simulations would be conducted: one with glazing transmittance = 0.1

and one with glazing transmittance = 0.9. A linear interpolation technique can be used to estimate window heat gain/loss for other values of glazing transmittance. In Step 4, a number of base building models generated from Step 1 to Step 3 would be used for Radiance and EnergyPlus simulations to calculate the daylighting and energy metrics, respectively.
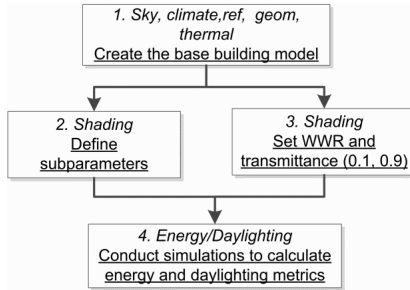
Figure 4. Process of generating pre-simulated data

Step 2 and part of Step 4 shown in Figure 4 are used to generate daylighting illuminance values for the different building models. The shading parameter in Step 2 includes sub-parameters (i.e., sensor point layout, window position and dimensions, glazing type, and overhang dimensions and location).

- Sensor point layout: The distance between the sensor points was set to 1m in both directions and the whole space was filled with sensor points.
- Window position and dimensions: The wall surface was divided into a number of small patches so that a combination of several consecutive patches could be grouped into a single window. The daylighting simulation may calculate the daylighting illuminance due to each separate patch and thus the overall daylighting illuminance of a window system is the sum of illuminances due to these separate patches.
- Glazing type: The glazing transmittance was set to 0.5. The illuminance changes in proportion to the change of glazing transmittance.
- Overhang dimensions and location: The daylighting system has two possible options: with overhang and without overhang. If the overhang is not installed, a single daylighting simulation using Radiance is capable of generating the illuminance from each patch. If the overhang is installed, a number of coefficients $C(oR, wW, wH)$ are generated first. Then the illuminance

  $E_p^o(oR, wW, wH)$ at the sensor point due to a single patch P can be calculated using Eq. (2), and the illuminance E at sensor points due to the whole window consisting of patches from m to n can be calculated by Eq. (3).

  $$E_p^o(oR, wW, wH) = C(oR, wW, wH) * E_p(oR, wW, wH) \qquad (2)$$

  $$E = \sum_{p=m}^{n} E_p^o(oR, wW, wH) \qquad (3)$$

Where oR = ratio of the overhang width to its length; wW = ratio of the distance between the center of patch p and the overhang edge (W_2) to the overhang

width (W_1) (i.e., $wW = W_2/W_1$ in Figure 5); wH= ratio of the vertical distance from the center of patch p to the overhang to the overhang width (i.e., $wH = H/W_1$ in Figure 5).
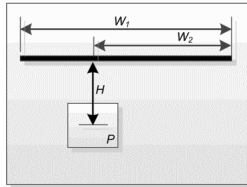
Figure 5. Elevation view of a patch and an overhang

Using Eqs. (2) and (3), pre-simulations would generate illuminance $E_p(oR, wW, wH)$ (i.e., illuminance due to each patch without overhang) and coefficient $C(oR, wW, wH)$. First, $E_p(oR, wW, wH)$ (without overhang) can be directly simulated by Radiance. Second, the calculation of $C(oR, wW, wH)$ is affected by the overhang dimensions and patch location. Thus, selection of different combinations of parameters oR, wW, wH would require a large number of simulations in order to calculate $C(oR, wW, wH)$. To reduce the large number of simulations, this research uses a simplified method to calculate $C(oR, wW, wH)$ by using a special window system (see Figure 6). The elevation view shows the 16 patches and an overhang represented by a solid black. There are four possible locations of the overhang as shown in Figure 6. For each overhang location, the width of an overhang is set to several different values. In this way, a number of different values of oR, wW, wH can be generated for each patch by changing the overhang location and dimensions. In addition, illuminances $E_p^o(oR, wW, wH)$ at the sensor points due to patch p will be calculated. Then $C(oR, wW, wH)$ can be calculated using Eq. (2). In the case shown in Figure 6, there are 16 $C(oR, wW, wH)$ values for each patch and for each possible overhang dimensions and location.
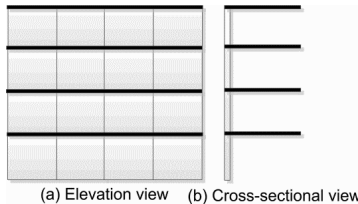
(a) Elevation view   (b) Cross-sectional view
Figure 6. Views of patches and four possible overhang locations

Expert system engine: Rule inference engine module

Three rules sets are used to extract information from knowledge database and conduct sensitivity analysis of the daylighting and energy performances of the user defined building model.

The first rule set is to select a base building model similar to the user defined building model (see Figure 7). The selected base building model should have the same climate zone, similar building interior reflectance and a similar building

location. If the building surface reflectance or the building location in the knowledge database is not the same as that in the user defined building model, a similar base building model with a minimal similarity index (S) should be selected. The similarity index is calculated by Eq. (4)

$$S = \sum (r_u - r_b)^2 + \Delta d / D \qquad (4)$$

Where r_u = interior reflectance of the user defined building model; r_b = interior reflectance of base building models in the knowledge; Δd = distance between the location of the base building model and the user defined building model; and D = mean value of the three closest locations.
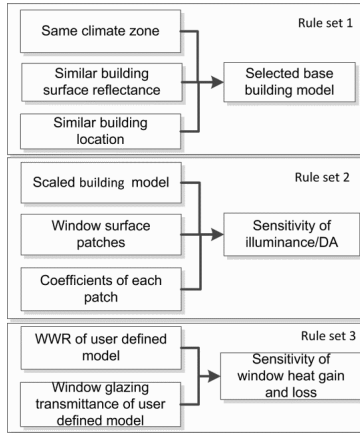
Figure 7. The three rule sets

After the based building model is selected by the rule set 1, the energy and daylighting data in Eq. (1) will be retrieved from the knowledge database. The purpose of the rule set 2 is to calculate daylighting and energy metrics based on the selected base building model. The base building model may have different actual dimensions. Therefore, the first condition of the rule set 2 is to scale the user defined building model to match the width of the base building model (see Figure 7). The second condition is to match the window surface area in the user defined model with the patches in the base building model. The third condition is the coefficients of each patch if an overhang is installed.

The purpose of the rule set 3 is to calculate the window heat gain/loss based on the base building model with a WWR value similar to the WWR of the user defined model. Thus, the first condition is WWR value of the user defined building model and the second condition is the glazing transmittance of the user defined model. The window glazing transmittance of the user defined model may not be equal to the glazing transmittance of the base building model. In this case, a linear interpolation technique can be used to derive the transmittances from 0.2 to 0.8 given the building model with transmittances 0.1 and 0.9.

The sensitivity data generated by the rule sets 2 and 3 are used to draw a 3D distribution figures of daylight illuminance, DA, and window heat gain/loss. These 3D figures may help designers to obtain a better understanding of the trend of the

change of the illuminance, DA and energy use if the inputted Menu item (5) in Figure 2 changes within a percentage specified in Menu item (6) in Figure 2. The final decision of how to optimize the design parameters is based on this information as well as designers' knowledge.

CONCLUSIONS

This research developed a framework of an expert system that can provide sensitivity analysis of the daylighting and energy performance metrics for different daylighting design parameters. When the design parameters fail to meet the performance requirement, the expert system can provide actions about how to improve performance. The expert system could provide a fast evaluation of daylighting system design parameters by utilizing the knowledge database.

Some limitations exist in this research. Only horizontal overhang was considered and thus other shading systems should be further analyzed. This research focused on the framework development of the expert system and implemented the user interface in the OpenStudio platform. The major part of the expert system has not been fully implemented yet. Thus, the future work includes implementing the full functionality of the expert system engine. The future research will focus on validating the accuracy of daylighting prediction, as well as conducting simulations to develop the knowledge database.

REFERENCES

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., and Glazer, J. (2001). "EnergyPlus: creating a new-generation building energy simulation program." *Energ Buildings*, 33(4), 319-331.
Gagne, J. M. L., Andersen, M., and Norford, L. K. (2011). "An interactive expert system for daylighting design exploration." *Build Environ*, 46(11), 2351-2364.
Gonzalez, A. J., and Dankel, D. D. (1993). *The engineering of knowledge-based systems: theory and practice*, Prentice Hall Englewood Cliffs, New Jersey.
Guo, B., Belcher, C., and Roddis, W. M. K. (1993). "RetroLite: An artificial intelligence tool for lighting energy-efficiency upgrade." *Energ Buildings*, 20(2), 115-120.
Hu, J., and Olbina, S. (2012). "Radiance-based Model for Optimal Selection of Window Systems." In: *ASCE 2012 International Conference on Computing in Civil Engineering*, Clearwater Beach, FL, USA.
Hunter, J., Dale, D., and Firing, E. (2012). "Matplotlib: python plotting.", <http://matplotlib.org/>, (Nov. 5, 2012).
Jung, T., Gross, M. D., and Do, E. Y. (2003). "Light pen–sketching light in 3d." *In Proc. of CAAD Futures*, Citeseer.
Nabil, A., and Mardaljevic, J. (2006). "Useful daylight illuminances: A replacement for daylight factors." *Energ Buildings*, 38(7), 905-913.
Reinhart, C. F., and Wienold, J. (2011). "The daylighting dashboard - A simulation-based design analysis for daylit spaces." *Build Environ*, 46(2), 386-396.

Providing Systems Engineering Perspective In a Capstone Project Setting to Monitor Performance of HVAC Systems

K. Knox,[1] A. Sanchez,[1] X. Liu,[2] S. Ergan[3] and B. Akinci[4]

[1]Master's Student, [2]Asset Manager, [3]Assistant R. Professor, [4]Professor, Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

ABSTRACT

Today's engineering problems require complex solutions from multiple disciplines. Hence, engineering classes should teach students to learn about how to solve multidisciplinary problems. This paper describes a project-based course that allows a group of students to learn and utilize knowledge and approaches from multiple disciplines, gain systems engineering perspective to solve a real-world problem. The objective of the project is to develop a system that can facilitate the monitoring and analysis of the performance of HVAC systems, which account for about 40% of total energy consumption in buildings. Students with various backgrounds worked together to develop a prototype system that automatically collects dynamic data from HVAC systems, and visualizes the real-time and historical data with floor plans using a web-browser-based interface. This prototype system also embeds fault detection and diagnosis (FDD) algorithms that frequently analyze the performance of HVAC systems. In developing such a system, the students had to learn and utilize knowledge in the domain of civil, mechanical, electrical engineering, computer science, and project management, approach the problem from a systems engineering perspective. This paper discusses the processes that the students went through to learn, implement and test the system in a real-world testbed.

INTRODUCTION AND BACKGROUND

According to the data from the U.S. Energy Information Administration (EIA), buildings use more than 42% of the total energy and are the largest energy consumer in the U.S. (EIA, 2008). Previous studies found that about 20-40 % of the energy used by the HVAC systems is wasted due to undetected faults, such as biased sensors, stuck dampers, and unstable controllers (Roth, 2005). The comfort and productivity of occupants are also impacted by the faults, if the indoor air quality, specifically temperature and concentration of CO_2 particles, is affected (Widodo, 2007). Hence, increasing energy efficiency of HVAC systems is essential to improvement of the overall energy performance of buildings.

Designing and operating HVAC systems involve engineers from a variety of disciplines, such as architecture, mechanical engineering and electrical engineering. Hence, efficiently operating the HVAC systems requires a multi-disciplinary approach, which is not commonplace in a classroom setting. There are various ways

to teach embedded sensing and topics related to performance monitoring of HVAC systems. One way is through a classical lecture-based setting, which teaches one general topic, e.g., sensing technology or thermal dynamics, in one course with greater depth. Another way is by merging these topics in a hands-on project in which the students can learn and practice multi-disciplinary knowledge using real-life projects. For such an approach, however students would need to have a systems engineering perspective, which require multidisciplinary approach to design and implement complex engineering systems through documentation of requirements, design synthesis and validation (INCOSE, 2004). The project described in this paper discusses a course designed to implement the second approach.

Carnegie Mellon University's Department of Civil and Environmental Engineering has established a graduate education program for its Masters students in the area of sensing, data processing and infrastructure management. A main capstone course is called AIS (Advanced Infrastructure Systems) project course, which is a team-based significant development project that is related to some area of infrastructure systems, industry driven, and built upon the knowledge, skills, and technologies learned in the core /specialist courses. An accompanying preparation course aims at starting the project and includes a set of lectures taught by experts in their areas to give the systems engineering perspective to students, and activities to start building the required background for the project.

This paper will provide; (a) an overview of a recent project that was provided to students to help them learn how to understand the performance monitoring of a new HVAC system and design a system for this purpose, (b) challenges associated with the current practice of HVAC performance monitoring, (c) the learning objectives (required skills) of the course and how the course was structured for students to achieve these learning objectives and successfully deliver the project, and (d) the lessons learned from this project to improve the course to better teach systems engineering perspective and the HVAC performance monitoring.

AN OVERVIEW OF THE COURSE PROJECT

During the Spring semester of 2012, a group of students enrolled in the AIS project course were assigned the task of developing a web-based interface to aid their client in the performance monitoring of a HVAC system installed in a recently renovated research laboratory. The objective of the project was to teach students embedding sensing, HVAC systems control, system analysis, fault detection and data modeling for storage of continuous sensing data and to help them approach the project using a systems engineering thinking.

Overview of the Testbed

This project used a secondary HVAC system installed in a research lab as the testbed. This secondary system has an air handling unit (AHU) that conditions three rooms in the lab. The AHU has a heating coil, a cooling coil, a supply fan and an economizer. The hot water and cold water are supplied from a plant in the campus. The lab included three rooms and each room is controlled by an individual variable air volume (VAV) box. Each VAV box contains a reheat coil and a damper. The entire system contains more than 50 sensors that measure the temperature, pressure and flow rates of the water and air. All sensor measurements and control signals are

communicated using BACnet (Building Automation and Control Network) protocol
and data is collected by the OSIsoft PI system.

Challenges in Monitoring the Performance of HVAC Systems

The following bullets discuss three challenges that inhibit system operators to
monitor and diagnose the performance of HVAC systems.

a. *Lack of systems to automatically collect dynamic data from the HVAC systems:*
HVAC systems continuously generate dynamic data, including sensor measurements
and control signals, which are critical for diagnosing the faults. Currently, only the
proprietary software developed by the vendors of HVAC systems manages dynamic
data. It is very difficult for the system operators to acquire the dynamic data for the
purpose of diagnosing faults.

b. *Difficulty in analyzing data within spatio-temporal contexts:* Analyzing the HVAC
system's performance requires information about the spatial environment and
dynamic conditions. For instance, simply reading the temperature sensor data in a
space does not support the analysis. However, if readings are compared with the
measurements in the adjacent spaces, it is possible to identify an abnormal
temperature control. Hence, the spatio-temporal contexts need to be integrated with
the dynamic data for performance analysis. Currently, it is challenging for the
operators to manually collect and integrate the contextual information.

c. *Having complex relationships between HVAC symptoms and faults:* Faults, such as
sensor drifts and duct leakage, are the problems associated with the HVAC
components and result in abnormal status or incorrect functioning. Symptoms are the
subset of the measurements and control signals that are impacted by faults. Symptoms
can be directly retrieved from the HVAC system and used to diagnose the faults.
However, there are many symptoms related to a given fault, making FDD quite
complex. The complex relationships between symptoms and faults make it very
challenging for manual diagnosis. This project required the students to implement and
test algorithms that can automatically detect and diagnose faults in HVAC systems.

The objective of this project has been to develop a system that addresses these
three challenges. Next section introduces the required skill sets and knowledge that
the team identified to develop such a system.

REQUIRED SKILL SETS AND KNOWLEDGE

In order to address these challenges and help system operators analyze the
performance of the HVAC systems, a wide range of knowledge and skill sets had to
be acquired throughout the semester to automatically collect, visualize and analyze
data from HVAC systems:

(1) Understanding the systems engineering perspective. Approaching this
interdisciplinary project required students to know systems engineering and
understand how to gather requirements, do design synthesis, verification and
validation to develop a system. Through an accompanying preparation course, the
students had these skill sets and implemented their knowledge in the project course.

(2) Understanding the basics of embedded sensing andHVAC systems control.
Many of the students had learned the sensing and data acquisition technologies from
another course, *Data Acquisition*. Also, in order to understand the HVAC control

system, students used a software application, called Learn HVAC (LearnHVAC, 2010).

(3) Understanding how to communicate with the HVAC systems to acquire sensor measurements and control signals. The students improved their programming capability, specifically in Java programming language, and learned the application interfaces provided by the PI system.

(4) Providing a web-browser-based user interface to visualize dynamic data from HVAC systems using spatial and temporal contexts. The students improved their knowledge on HTML, PHP and JavaScript programming language to develop the web-browser-based user interface.

(5) Embedding algorithms that automatically detect and diagnose faults in HVAC systems. The students reviewed several scientific publications to learn and implement the FDD approaches that have been developed by researchers.

Collaborative Team

Due to the interdisciplinary nature of the project, a collaborative team needed to evolve from what initially appeared to be a homogenous group of engineering students. Even though graduate students in the Department of Civil Engineering undertook the project, they brought unique backgrounds from prior fields of study. Undergraduate degrees represented by the project team included not only Civil Engineering, but also Electrical Engineering and Physics backgrounds.

Useful Tools and Techniques

The AIS program offers courses that allow students to supplement their knowledge of both technical and management related topics, such as Data Acquisition, Data Management, Computer-based Approaches for Search and Decision Support and Project Management for Construction. In addition, the project preparation course enabled the students to learn from a number of lectures and modules that were implemented during the actual project course. For example, requirements elicitation is the process of discovering the needs of the users and stakeholders of a system or service, which in this case is a client enacting the role of a facility manager with knowledge of HVAC systems, building automation software and FDD algorithms. Another technique that the group learned how to implement throughout the project was rapid prototyping. In the context of the design and planning phase of any project, rapid prototyping is the iterative process involving four basic phases that are connected in a cyclical pattern. These four phases are (1) requirements elicitation, (2) design, (3) implementation, and (4) testing. By using rapid prototyping method, the group was able to start with a simple, basic test case or product design and then expand to greater levels of detail with subsequent iterations.

PROCESS AND APPROACH

Overview of the Project Delivery

Figure 1 shows an overview of the processes involved during project preparation and project courses and the transitions made between learning about the project and its implementation. As previously discussed, the preparation course's primary focus was to learn the systems engineering phases and find information about the components of the project to consider in the second semester project course.

Identifying requirements, developing the system architecture, prototyping, testing and structuring the group were main processes iterated throughout the project. Requirements elicitation with an expert in building performance monitoring led the group to determine a web-based interface to be the preferred product to address the current challenges faced by facility managers.
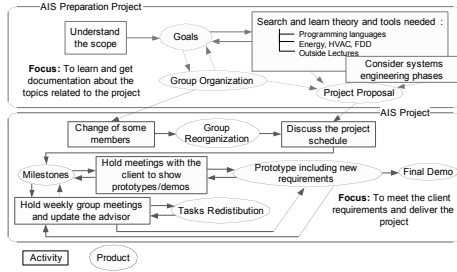
Figure 1. Overview of the workflow for the project.

System Architecture Design

This particular project was developed into a three-tier architecture. The tiers included in such a model are data, analysis, and visualization. The data layer is responsible for managing the data collected from the HVAC system sensors (Figure 2). In this case the Building Automation and Control Network (BACnet) protocol was used to access the dynamic sensor information from the HVAC system (BACnet, 2010). The analysis layer consists of equations that perform an analysis of the data according to a rule-based FDD approach, which will be discussed later on. The visualization layer consists of the web pages, which the client has direct access and interaction with.
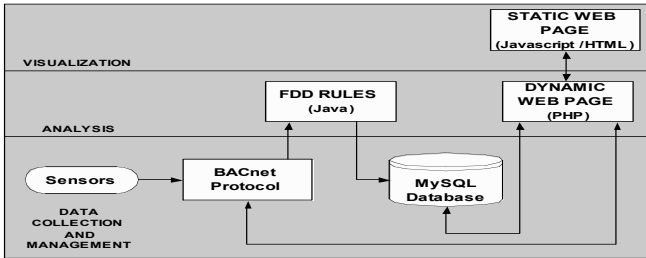
Figure 2. Three-tier architecture design.

Project Decomposition

The web interface developed for the clients to monitor the performance of the HVAC system from locations outside the lab, displays both real-time and historic sensor data, visualizes comparisons between sensors, and stores information about any faults detected in the system. The group was initially divided into three subgroups: HVAC systems and FDD, Database/ Data collection, and User Interface.

The first group was in charge of understanding the proper and expected performance of the HVAC system. Additionally, this group needed to understand the

way the sensed variables interacted, so they could help implement the rules to some already developed FDD algorithms. The two methods chosen for FDD analysis were the rule-based approach and the historical-based approach (Katipamula, 2005). The rule-based approach utilizes limits and alarms to alert a possible fault while the historical-based approach forms a statistical model of faults that have occurred over time. The database and data collection group had to collect information and create a database. This database saved the data regarding to the FDD approaches that were evaluated. The last group was responsible for the development of the web interface that displays information in a series of pages, or modules including the system status, room and AHU, and FDD modules discussed below.

System Status Module

This module serves as the primary connection between the other user pages. It also shows the temperature and set point for each room in real time. It provides an overview of the conditions of the three rooms in the lab for facilities managers.
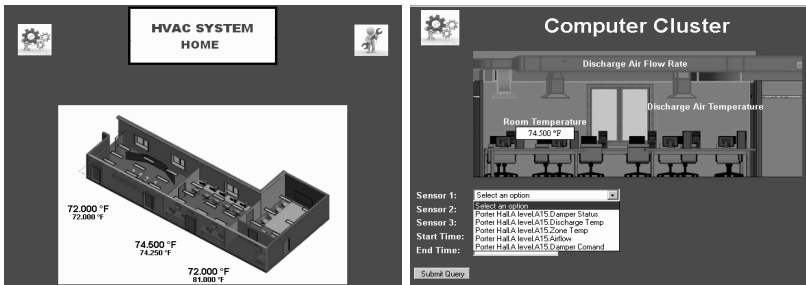
Figure 3. The user interface for system status and room modules.

Room and AHU modules

For each room in the lab, there is a module and interface (Figure 3) that shows an image of it with the real time value of three variables of the system: *Discharge Air Flow Rate, Discharge Air Temperature* and *Room Temperature*, these data points are used to measure and control temperature for each room. The text box, where the value is shown, appear when the user hovers over a particular sensor name of a 3D image.
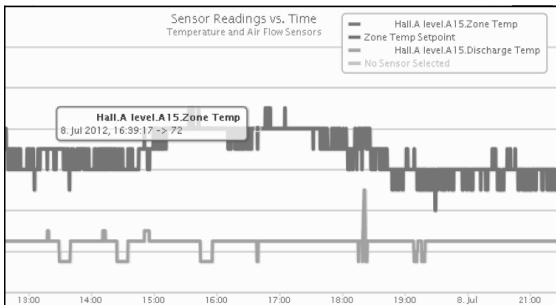
Figure 4. Graphing feature of the interface.

Another feature on each room module and user interface is the ability to query historical data from the PI server by selecting a sensor from a drop down menu up to

three different variables. Figure 4 shows an example of the sensor data plotted as a function of time. Each data series is plotted with their own y-axis that is color coordinated with its respective data series for clarity.

The AHU module and interface has almost the same features of the rooms' modules, but for the historical data graph, it has the capability to handle up to five variables of the sensors related to the AHU. Additionally, real-time values of all the related/selected variables are shown in a table.

FDD module

In this module and user interface, a summary of the faults detected in the HVAC system is presented. It is divided into four sections that correspond to each one of the rules that are being evaluated: *Air Temperature Sensor Fault*, *Airflow Sensor Fault*, *Cooling Coil Temperature Sensor Fault* and *Operation Fault*. This classification is done when the information is pulled from the MySQL database. The faults showed are being filtered by time. Information associated with each fault includes the date and time of the detected fault and a record of fault acknowledgment.

LESSONS LEARNED

The group learned the following lessons related to the value of a project based educational environment from their experience in systems engineering:

(1) *Approach the project using a systems engineering process:* The group benefited from the steps of the systems engineering process (requirements identification and documentation, design synthesis, verification, validation and testing) and were able to decompose the responsibilities accordingly.

(2) *Learn about HVAC controls and embedded sensing systems through interactive resources and with hands-on in a real system:* Research regarding the correct performance of HVAC systems was not sufficient with just a literature review. The group had several trips to the mechanical room where the HVAC system was primarily located and the client explained how the sensors were integrated into the system according to the BACnet protocol.

(3) *Use an iterative process:* The group learned the benefits of using iterative processes, such as rapid prototyping. Working in a circular manner as opposed to a linear, "waterfall" type process, enabled the group to fix problems in the system before they became too difficult or complex to correct.

(4) *Keep communication between group members flowing, use a communication medium:* The group learned that efficient communication among group members, keeping a schedule of the deliverables by subgroup members and controlling this schedule were necessary for success. In order to frequently share updates and feedback from both the client and the course advisor, the group relied on not only the project manager, but also an online tool.

CONCLUSIONS

The AIS preparation and project courses give students a chance to work on a real-life problem in a real-world facility. The lessons learned from this experience,

such as how to gain familiarity with and knowledge of an embedded sensing environment, can be transferred to projects encountered in an engineering career. The multi-disciplinary and collaborative team that the students created simulates the type of environment students will encounter in their later professions.

Students realized that limited background knowledge in a particular discipline should not be an obstacle to participating in a given project. The learning modules and lectures that are an integral part of the preparation course equipped the group with useful tools such as requirements elicitation and systems engineering.

In conclusion, the hands-on project course setting gave graduate students the opportunity to acquire the necessary skills and knowledge to create a HVAC performance monitoring system based on embedded sensors. The group was able to overcome a number of challenges and successfully collected HVAC sensor data, performed FDD analysis, and presented the results in a web-based user interface.

ACKNOWLEDGEMENTS

REFERENCES

EIA. (2008). The 2007 Commercial Building Energy Consumption Survey (CBECS). U.S. Energy Information Administration: Washington, D.C.

ASHRAE, A., ASHRAE/IESNA Standard 90.1-2004, Energy Standard for Buildings Except Low-Rise Residential Buildings. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc., Atlanta, Georgia, USA, 2004.

Roth, K.W., et al. (2005). "The Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential." TIAX LLC: Cambridge, MA.

Widodo, A. and B.S. Yang. (2007). "Support vector machine in machine condition monitoring and fault diagnosis." Mechanical Systems and Signal Processing, 21(6), 2560-2574.

OSIsoft. (2012). Official website for PI system from OSIsoft. [cited 2012 July 31, 2012]; Available from: http://www.osisoft.com/.

LearnHVAC. (2010). Learn HVAC official website. Available from: http://www.learnhvac.org/.

Katipamula, S. and M. Brambley. (2005). "Methods for fault detection, diagnostics, and prognostics for building systems-A review, part I." HVAC&R Research, 11(1), 3-25.

Epistemic Modeling for Sustainability Knowledge Management in Construction

Lu Zhang[1] and Nora M. El-Gohary, A.M.ASCE[2]

[1]Graduate Student, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, IL 61801; FAX (217) 265-8039; email: luzhang7@illinois.edu
[2]Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, IL 61801; PH (217) 333-6620; FAX (217) 265-8039; email: gohary@illinois.edu

ABSTRACT

One of the key factors for enabling sustainable construction is facilitating the acquisition, transfer, and exchange of knowledge about sustainable construction practices (SCPs). However, knowledge about SCPs is complex, interdisciplinary, multi-faceted, and context-sensitive. Semantic modeling is a key to achieving context-awareness. The development of a context-aware knowledge management (KM) system can be successfully achieved based on a semantic (computer-understandable and meaning-rich) model. This paper proposes an epistemic model for SCPs based on epistemology (the theory of knowledge and knowing). The proposed epistemic model is a theory-based, semantic model for representing and reasoning – in a context-aware manner – about: 1) the knowledge of SCPs and 2) the process of knowing this knowledge. This paper starts by presenting our analysis of the requirements of sustainability KM in the construction domain and the limitations of existing methods and systems. It follows by presenting our modeling approach for KM system development, which is epistemology-based, context-aware, and semantic. Finally, the paper presents our initial modeling efforts towards a formal epistemology for SCPs and discusses its application as the backbone of our proposed KM system.

INTRODUCTION

Construction practices performed during the construction phase may significantly affect – positively or negatively – the society, economy, and the environment (Son *et al.* 2011). For example, construction practices may consume large amounts of water, energy, and natural resources; and may cause significant negative impacts, such as waste generation, air and water pollution, noise, health and safety problems, community disruption, etc. To alleviate negative impacts and encourage positive impacts, contractors are required to implement sustainable construction practices (SCPs) during the construction phase (CII 2009). Nowadays, the awareness and understanding of the importance of implementing SCPs during the construction phase is greater than that of the past (Son *et al.* 2011). There is also a strong realization that a knowledge-intensive mode is needed to promote the achievement of sustainability in the construction industry (Khalfan *et al.* 2002). As

202

part of that, there is a stronger need for allowing project participants (e.g. contractors) to acquire, transfer, and exchange knowledge about SCPs to enable better selection and implementation of SCPs during the construction phase.

However, the acquisition, transfer, and exchange of knowledge about SCPs – for example which SCP to implement and how to implement it – is not an easy task, because SCPs are highly context-sensitive and the choice of which SCP to implement and how to implement it are, similarly, context-dependent (e.g. dependent on project type, project size, project location, regulatory requirements, site characteristics, etc.). In a similar way, the process of knowing about SCPs is context-sensitive; it is sensitive to the context of the knower (e.g. knower preference, knower background), the knowing process (e.g. knowing location, knowing device), and SCPs knowledge (e.g. SCPs implementation context). A new approach for the development of a context-aware KM system is, thus, crucial to enable the implementation of SCPs by facilitating context-aware acquisition, transfer, and exchange of SCPs-related knowledge within the domain.

CURRENT EFFORTS AND LIMITATIONS OF KNOWLEDGE MANAGEMENT IN THE CONSTRUCTION DOMAIN

Research on KM has been growing and expanding rapidly in a variety of fields in recent years (Rezgui *et al.* 2010). The concept of KM is broad. Scarbrough *et al*. (1999) defines KM as "any process or practice of creating, acquiring, capturing, sharing and using knowledge, whenever it resides, to enhance learning and performance in organizations". In terms of KM in the construction domain, a variety of studies have been conducted, and a set of practical guidance, approaches, and systems have been developed. Examples include the Constructability Implementation Program led by CII (1993) – a program for capturing and sharing of constructability knowledge, CLEVER (Kamara *et al.* 2002) – a framework for knowledge transfer in multi-project environment, Dr. Check (Soibelman *et al.* 2003) – a system for incorporating personal experiences and lessons learned into corporate knowledge, and KLICON (McCarthy *et al.* 2000) – a project for providing an understanding of how knowledge is gained and how learning can be formalized. Efforts focusing on the management of sustainability knowledge in the construction domain include 1) C-SanD project (Wetherill *et al.* 2007), which uses KM and information systems to facilitate the creation, capture, and transfer of sustainability knowledge in the construction domain; 2) SMAZ (Shelborun *et al.* 2006), which maps sustainability issues onto a generic project process to identify sustainability actions needed at different stages of a project lifecycle; and 3) Classification System for Capital Project Offices (CPO) (Tan *et al.* 2012), which assists CPO organizations to store and manage their sustainability knowledge.

Collectively, existing research and system development efforts in the area of KM provide important and valuable work towards effective management of sustainability knowledge. However, in our analysis, these efforts have four main limitations: they 1) place more emphasis on sustainable design practices with inadequate attention to sustainability practices that can be implemented during the construction phase; 2) lack an epistemological foundation for KM; 3) offer tools that lack adequate awareness of and adaptation to the characteristics of the knower, the types of knowledge, and the circumstances of the knowing process; and 4) focus on

the discussion of methods or tools needed for promoting KM for sustainable construction, but without offering a detailed-enough knowledge model that is needed to support day-to-day knowing (e.g. search, retrieval, acquisition, etc.) of sustainability knowledge.

PROPOSED MODELING APPROACH FOR KNOWLEDGE MANAGEMENT SYSTEM DEVELOPMENT

This paper proposes a modeling approach for KM system development that aims at facilitating effective, context-aware, and meaning-rich management of sustainability knowledge during the construction phase. This approach: 1) utilizes epistemology (theory of knowledge and knowing) as a theoretical, epistemological foundation for KM; 2) focuses on context-awareness and proposes a wider, multi-faceted perspective in modeling the concept of "context"; and 3) employs a formal semantic (computer-understandable and meaning-rich) model to represent the knowledge of SCPs and the process of knowing this knowledge.

EPISTEMIC APPROACH

Our proposed modeling approach is based on epistemology – theory of knowledge and knowing. Epistemology is the branch of philosophy that deals with the knowledge nature and scope of knowledge (Muis 2004). The term "epistemology" is derived from the Greek word "epistēmē" meaning "knowledge and science", and "logos" meaning "study of" (Honderich 1995). It attempts to answer the questions: What are the types of knowledge? What is the structure of knowledge? What are the sources of knowledge? How is knowledge acquired? What are the necessary and sufficient conditions of knowledge? How are knowledge claims justified? How does knowledge flow (Steup 2011)?

In modeling a KM system, "it is necessary to understand the broad epistemological spectrum that can enable effective utilization of computerized systems for knowledge management" (Jayatilaka and Lee 2003). In order to better manage knowledge, one has to understand what the nature of knowledge is, what the sources of knowledge are, and how knowledge can be acquired, etc. Epistemology provides a rationale for managing knowledge, defining the process of managing knowledge, and enabling the acquisition of knowledge by giving sufficient answers to these questions. The investigation of the knowing process, the factors affecting the knowing process, the knowers, etc., enhances the analysis of system requirements and facilitates the understanding of how knowers can effectively and efficiently acquire, transfer, and exchange knowledge using that system. It is through this type of integrative understanding and modeling that a successful KM system can be designed and developed (Chun *et al.* 2008).

CONTEXT-AWARE APPROACH

Industry professionals (a type of knowers) consistently report difficulties in efficiently finding and accessing the information/knowledge they need – information/knowledge that is relevant and specific to their needs (TRB 2006). "Lack of such access to appropriate information may often result in less effective decisions, duplication of effort, and greater cost" (Harder and Tucker 2003). A context-aware modeling approach would facilitate more relevant and specific knowledge acquisition by adapting to the various contexts of the knower, the knowing process, and the

knowledge. This is essential since the process of knowing is highly context-dependent: 1) the knowing process can be conducted in a variety of knowing locations, be affected by the knowing environment (e.g. noise, weather, etc.), and be performed using various knowing devices (e.g. laptop, smartphone); 2) every knower who is involved in the knowing process may have different interests or preferences which may vary by age, gender, academic level, etc.; and 3) the same piece of knowledge may have different semantics in different contexts. In terms of acquisition of SCPs knowledge in particular, 1) projects have different contexts in terms of location, type, size, budget, schedule, etc.; and 2) construction sites have dynamic contexts in terms of terrain, soil, water, air, weather, etc.

SEMANTIC APPROACH

Semantic modeling is a key enabler in our proposed approach, in three main ways. First, semantic modeling represents the epistemology (theory of knowledge and knowing) of SCPs in a formal and semantically-rich way, thereby serving as a backbone for the development of an epistemologically-grounded KM system. Second, semantic modeling is a key for facilitating context-awareness (Yilmaz and Erdur 2012) by offering a way to formally model the concept of 'context', its subconcepts, and its interrelationships with other epistemic concepts (e.g. knowing process). Third, semantic modeling can provide a domain-specific, unambiguous, and formalized representation of SCPs knowledge and the relations 1) between the knowing process and SCPs knowledge, and 2) between the contexts and SCPs knowledge.

INITIAL EPISTEMIC MODEL FOR SUPPORTING KNOWLEDGE MANAGEMENT IN CONSTRUCTION

Our proposed epistemic model is a formal, theory-based, meaning-rich, and domain-specific semantic model. It is composed of concepts, inter-concept relations, and axioms. Concepts represent the "things" that describe the knowledge of SCPs and the process of knowing this knowledge in a context-aware manner. Relations define the interactions between the different concepts. Axioms 1) specify the definitions of concepts and relations, and constraints on their interpretation; and 2) define the rules and requirements for context-aware knowing and the implementation of SCPs.

A preliminary upper-level epistemic model, showing the most abstract epistemic concepts of the model, is depicted in Figure 1. A "knowing process" *has* a "knower" *involved*, *is conducted in* a "knowing context", *aims at* a "knowing objective", *results in* a "knowing outcome", *uses* a "knowing technique" and a "knowing resource", and *is constrained by* a "knowing constraint". A "knower" *acquires* "knowledge". "Knowledge" *has* a "knowledge attribute" and a "knowledge modality", *belongs to* a "knowledge family", *comes from* a "knowledge source", and *can be represented in* a "knowledge item". "Epistemic context" and "knowing technique" *directly affect* the "knowing process", and *indirectly affect* the "knowing outcome". "Knowledge" is any information, fact, description, or skill acquired through experience or education and processed in human mind. A "knowing process" is a process of acquiring, transferring, and exchanging "knowledge". A "knower" is a person or a group of persons who is/are involved in the process of knowing. A "knowing objective" is a goal that the "knowing process" aims at achieving or will be able to achieve. A "knowing outcome" is a final result that a "knower" acquires or

should be able to acquire through "knowing process". A "knowing technique" is any method, means, or way that is adopted to know certain "knowledge". A "knowing resource" is a source of supply, support, or aid to the process of knowing. A "knowing constraint" is a limitation or restriction that constrains the "knowing process". A "knowledge attribute" is a characteristic that can define or describe "knowledge". A "knowledge modality" is a characteristic that defines the belonging criteria to a "knowledge family". A "knowledge family" is a group or classification of knowledge. A "knowledge item" is a physical repository that stores such "knowledge" or is a symbolic manifestation of the "knowledge". A "knowledge source" is an essential origin or prime cause of "knowledge". An "epistemic context" is any situation, setting, environment, or set of parameters that affects the knower's knowing process directly and affects the "knowing outcome" indirectly.
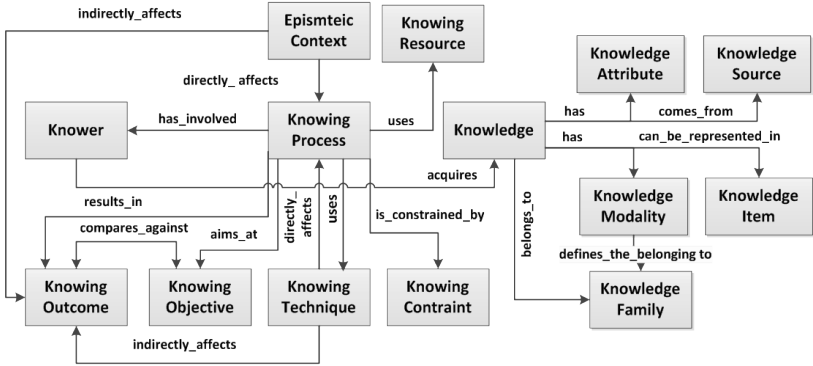
Figure 1. Preliminary upper-level epistemic model.

The most abstract concepts (shown in Figure 1) have a set of sub-concepts forming a concept hierarchy. For space limitations, the full concept hierarchy is not presented in this paper. But, as a partial example, the upper-level part of the preliminary 'epistemic context' hierarchy is presented in Figure 2. An "epistemic context" is a "knower context", a "knowing context", or a "knowledge context". A "knower context" is a specific interest, preference, or personal profile that a "knower" has. Different "knowers" may have different "knower contexts". A "knowing context" is a background, environment, setting, or situation where a "knowing process" occurs,
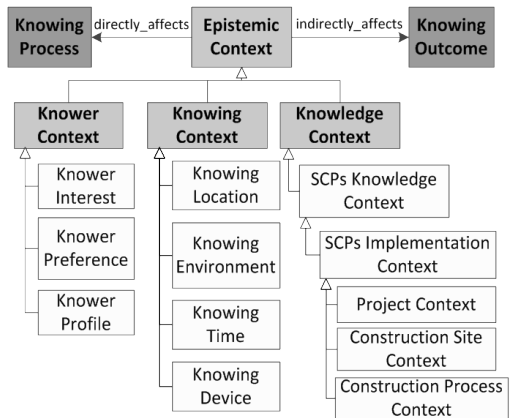
Figure 2. Partial epistemic context hierarchy.

such as a "knowing location", a "knowing environment", a "knowing time", or a "knowing device". A "knowledge context" is a collection of relevant conditions and surrounding influences that make the semantics or function of the knowledge unique and comprehensible to that condition. Knowledge context supplements knower context and knowing context with domain-specific relevance related to the types of knowledge. Since our scope is limited to SCPs knowledge, our model focuses on describing the "SCPs knowledge context" (a type of "knowledge context"); for example, describing its implementation context in terms of "project context", "construction site context", and "construction sequence context".

APPLICATION OF THE EPISTEMIC MODEL IN A KNOWLEDGE MANAGEMENT SYSTEM

Our proposed epistemic model will serve as a foundation for developing a KM system that can support context-aware acquisition, transfer, and exchange of SCPs knowledge within the construction domain. The proposed KM system consists of four main modules, as illustrated in Figure 3: search and retrieval module, summarization module, recommendation module, and classification module. For search and retrieval, we are taking a federated search approach, which enables users to search multiple sources of data that are related to SCPs knowledge through one single query. For such federated search, data remain in their sources, while our KM system serves as an integrator to

Figure 3. Main modules of our proposed KM system.

federate and search all these sources. Users would be able to view all retrieved data in one single list and link directly to their sources. In our KM system, data sources include publicly available online databases, online journals, web pages, etc. from researchers, government agencies (e.g. EPA, CADOT, etc.), non-profit organizations or consortiums (e.g. CII, USGBC), and professional societies (e.g. ASCE), etc.

For better illustration of how the epistemic model would support the search and retrieval module, an example use case scenario for context-aware semantic search and retrieval is illustrated in Figure 4. The example deals with searching and retrieving knowledge about silt fence as a sediment control practice. A construction engineer who works for a project in California wants to know about the installation method of a sediment control practice that captures sediment that 1) runs off from a slope and 2) caused by storm water. The KM system being aware of the project context, construction site context, knower context, knowing context, and domain knowledge, can automatically extract relevant and specific knowledge about the installation method of silt fence, applicable to the subject contexts. The knowledge is

extracted from the California Department of Transportation (2003) Construction Site BMP Field Manual and Troubleshooting Guide. This search aims at improving the search accuracy by addressing the specific needs or requirements of acquiring SCPs related knowledge, as well as implementing SCPs.
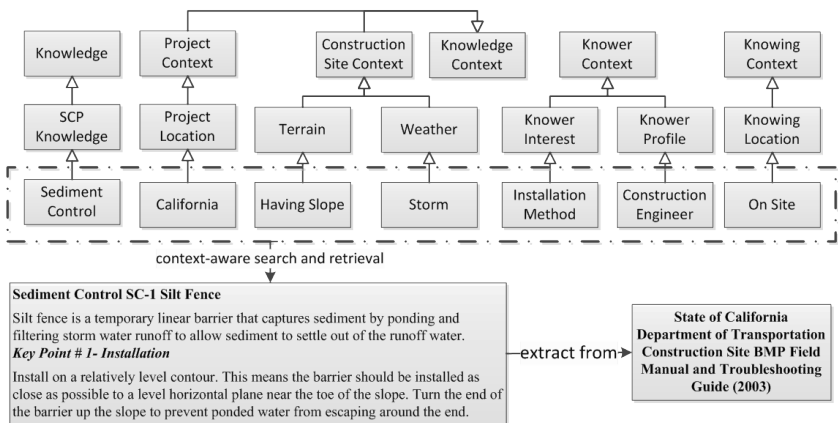
Figure 4. Context-aware search and retrieval in the proposed KM system: an example use case scenario.

CONCLUSIONS AND FUTURE WORK

This paper proposed a modeling approach for knowledge management (KM) system development that aims at facilitating effective, context-aware, and meaning-rich acquisition, transfer, and exchange of sustainability knowledge during the construction phase. Towards developing the proposed KM system, an epistemic model (semantic model based on epistemology (the theory of knowledge and knowing)) for supporting context-aware KM in the constructing domain has been developed and presented in this paper. The proposed model will serve as a foundation for developing the proposed KM system. It offers a leading effort in the area of formal epistemic modeling and application in the sustainable construction domain. Future/ongoing research by the authors will focus on evaluating the epistemic model, implementing the model within the proposed KM system, and evaluating the performance of the various modules (e.g. search and retrieval) of the system.

REFERENCES

California Department of Transportation (CADOT). (2003). "Construction Site Best Management Practice Field Manual and Troubleshooting Guide." *CTSW-RT-02-007*, California Department of Transportation, Los Angeles, CA.

Construction Industry Institute. (CII). (2009). "Sustainable design and construction for industrial construction: Implementation resources." *Implementation Resource 250-3*, CII, Institute, Austin, TX.

Chun, M., Sohn,K., Arling,P., and Granados,N.F.(2008). "Systems Theroy and Knowledge Management Systems: The Case of Pratt-Whitney Rocketdyne."

Proc., 41st Hawaii International Conference on System Sciences, IEEE, Washington, D.C., 336-346.

Harder, B.T. and Tucker, S.L. (2003). " Scoping Study for a National Strategic Plan for Transportation Information Management." *Report NCHRP/20-7/142*, Transportation Research Board, Washington, D.C.

Honderich, T. (ed.). (1995). *The Oxford Companion to Philosophy,* Oxford University Press, New York.

Kamara, J.M., Anumba, C.J., and Carrillo, P.M. (2002). "A CLEVER approach to selecting a knowledge management strategy." *Int. J. Proj. Manage*., 20, 205–211.

Khalfan, M. M. A., Bouchlaghem, D. M, Anumba, C. J. and Carrillo, P. M. (2002). "A framework for managing sustainability knowledge the C-SAND approach." e-Sm@rt, Salford, UK 19-21.

Jayatilaka, B. and Lee, J. (2003). "An Epistemological Taxonomy for Knowledge Management Systems Analysis." *Proc., 36th Hawaii International Conference on System Sciences*, IEEE, Washington, D.C., 10-20.

McCarthy, T.J., Kahn, H.J., Elhag, T.M.S., Williams, A.R., Milburn, R., Patel, M.B. (2000). "Knowledge management in the designer/constructor interface." *Proc., 8th International Conference on Computing in Civil and Building Engineering*, Reston, VA, 836–843.

Muis, K. R. (2004). "Personal epistemology and mathematics: A critical review and synthesis of research." *Rev. Edu. Research*, 74, 317–377.

Rezgui,Y., Hopfe, C.J. and Vorakulpipat, C.(2010). "Generations of knowledge management in the architecture, engineering and construction industry: An evolutionary perspective." *Advan. Eng. Infor*., 24 , 219-228.

Scarbrough, H., Swan, J. and Preston, J. (1999). *Knowledge Management: A Literature Review*. Institute of Personnel and Development, London, UK.

Soibelman, L., Liu, L., Kirby, J., East, W., Caldas, C., and Lin, K. (2003). "Design review checking system with corporate lessons learned." *J. Constr. Eng. Manage.*, 129(5), 475–484.

Son, H., Kim, C., Chong, W. K., Chou, J. (2011). "Implementing sustainable development in the construction industry: Constructors' perspectives in the US and Korea." *Sustain. Development*, 19, 337–347.

Steup, M. (2011). "Epistemology", *The Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/archives/win2011/entries/epistemology/>. (May 18, 2012).

Tan, X., Lin, K., Anumba, C. J., Tan, X., and Li, H. (2012). "Exploratory Study of Sustainability Knowledge Management in Capital Project Offices." *Construction Research Congress*, West Lafayette, IN, 1809-1819.

Transportation Research Board (TRB). (2006). "Transportation knowledge networks: a management strategy for the 21st century." *TRB Special Report 284*, TRB, Washington, D.C.

Wetherill, M., Rezgui, Y., Boddy, S., and Cooper, G. (2007). "Intra and inter-organisational knowledge services to promote informed sustainability practices." *Compu. Civ. Eng*., 21, 78–79.

Yilmaz, O. and Erdur, R.C. (2012). "iConAwa- An intelligent context-aware system." *Expert Syst. Applic.*, 39, 2907-2918.

Lessons learned from developing immersive virtual mock-ups to support energy efficient retrofit decision making

Xue Yang[1], Yifan Liu[2], Semiha Ergan[3], Burcu Akinci[4], Robert M. Leicht[5] and John I. Messner[6]

[1]PhD Student, Asst. R. Professor[3], Professor[4] Dept. of Civil & Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA, USA, 15213; email: xueyang@andrew.cmu.edu, semiha@cmu.edu, bakinci@cmu.edu
[2]PhD Student, Asst. Professor[5], Professor[6], Dept. of Architectural Engineering, Penn State, University Park, PA, USA; email: yxl5209@psu.edu, rmleicht@engr.psu.edu, jmessner@engr.psu.edu

ABSTRACT

For energy efficient retrofit projects, interactions of integrated design teams have a significant impact on the life-cycle energy saving decisions. While evaluating design options, the team must consider results from various simulation tools (e.g., energy, day-lighting) with the context of the 3D model. However, traditional communication creates a barrier for integrated design teams to make decisions efficiently as information is frequently conveyed through documents in tabular formats, and perceived outside of the context of the 3D representation. One of the primary needs of the team is to integrate the assessment of options with simulation results. This research focuses on using immersive visualization to support multi-stakeholder decision-making processes for advanced energy retrofit projects. An immersive virtual mock-up was developed for visualizing energy modeling assumptions and energy efficient metrics to facilitate integrated design team interactions toward evaluating retrofit alternatives. The mock-up incorporates and enables integrated visualization of a 3D model from a Building Information Model (BIM) application and engineered data from simulation tools, and was developed and compared using a commercial immersive visualization environment and a real-time 3D game engine. This paper provides a comparison of the process for developing the virtual mock-up in two different environments, and lessons learned including the challenges related to exchanging semantic information and developing reusable modules for visualizing simulation results in virtual environments. These lessons will inform the development of a repeatable process that can be scaled to different types of buildings and design decisions.

INTRODUCTION

Studies have shown that buildings account for 41% of the total energy consumption in the US (DoE 2008). With the degradation of old building systems and development of new technologies, there are huge opportunities for energy savings by applying energy efficient retrofits to existing buildings. The advanced energy retrofit process is a multi-stakeholder (including architects, engineers, facilities managers,

 owners, etc.) decision-making process, which includes assessment and upgrading/replacement of existing mechanical systems, envelope, and other systems, in an integrated manner to improve the energy performance of existing buildings. Integrated design team decisions play an important role in the process and the teams' final decisions have significant impacts on the retrofit and lifecycle energy savings. Evidence suggests that good design practice can save up to 50% of building energy use, while poor practice can increase building energy use by 60-210% (Heller et al. 2011). However, many factors, such as climate, cost, energy saving, occupancy, and equipment, among others, need to be taken into consideration when developing energy saving solutions, and a large amount of design and engineered data is generated in the process. The amount of information can become overwhelming for the design team and hinder effective decision-making. During retrofit design review meetings, every party needs to convey essentials of the simulation models (i.e., the assumptions and results) and relate them to the design options reviewed by the design team members.

Virtual mock-ups using immersive projection displays have been shown to help project teams gain a better understand of a building design, and promote conversations that are valuable for problem solving (Messner 2006). To facilitate the integrated design teams to better evaluate retrofit alternatives, the authors are using advanced visualization technology to improve information exchange and communication by embedding semantic information for energy retrofit alternative evaluation into a virtual mock-up. This paper describes the process of developing the virtual mock-up in two software environments, the lessons learned along the way, and the analysis of the strengths and limitations of the environments that were used in this research in supporting semantic information visualization by using immersive virtual reality. The lessons learned are being applied to the development a repeatable process for rapidly developing virtual mock-ups with multiple sources of data for a building to support different types of design decisions.

BACKGROUND RESEARCH AND PROBLEM STATEMENT

The Architecture Engineering and Construction (AEC) Industry has its intrinsic characteristic of complexity in terms of the building product as well as the process of designing and constructing it. Prototyping, such as creating a virtual mock-up, is often very helpful in analyzing and addressing potential issues before the implementation of a project (Gopinath & Messner 2004). Using virtual prototyping tools, one can create a computer-based model to graphically simulate and demonstrate a building product. Development of virtual mock-ups, which are capable of demonstrating building product and process models in a realistic manner at low cost, has gained attention in academia as well as in the industry in recent years. Furthermore, immersive virtual mock-ups have additional advantages of adding a sense of realism, enabling 1:1 scale display and involvement of multi-stakeholders (Gopinath & Messner 2004, Leicht et al. 2010). They have also been proven to be a great alternative or means of augmenting physical prototypes (Maldovan et al. 2006). Many studies have been conducted to use immersive virtual mock-ups to support decision-making for design review (Leicht et al. 2010, Dunston et al. 2007) and

construction schedule simulation (Whisker et al. 2003, Messner 2006). These studies focused on visualizing and rendering 3D building product models or 4D construction processes realistically in immersive settings, but generally were limited in displaying a wide variety of semantic information other than construction scheduling, and had limited permission for user interactions other than navigation or selective viewing of 3D representations of design elements.

The work presented in this paper is a practical implementation of Virtual Reality (VR), but augments the virtual mock-up with real energy related semantic information. In the human-computer-interaction (HCI) domain, there are research studies exploring integrating virtual environments with information visualization: yielding information-rich virtual environment that enhance the virtual environment with the addition of related semantic information (Bowman et al. 2003). The virtual mock-up developed in this research can also be defined as information-rich virtual environment.

Though virtual reality has been used for various tasks in the AEC domain, information-rich virtual environments have not been widely explored in the domain. This research includes an experiment focused on tailoring both a commercial VR software package and a game engine to develop virtual mock-ups that integrate 3D building models and related semantic information. This paper provides an overview of the process for developing the virtual mock-ups using the two different tools, and lessons learned in the process, including a discussion of the challenges encountered, and analysis of the strengths and limitations of the two environments.
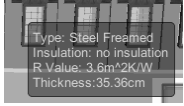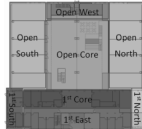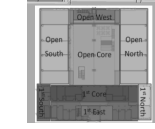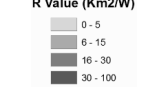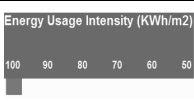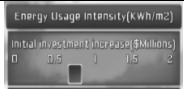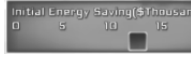
RESEACH METHOD

The objective of this study was to explore the development of semantically rich virtual mock-ups using two different environments in a case study and to summarize the lessons learned to streamline development of virtual mock-ups in different projects with similar functional requirements. For this purpose, the authors used two environments to develop the same virtual mock-up: one is a typical virtual reality (VR) development environment, and the other one is a 3D game engine. The process for generating the virtual mock-up includes several steps. First, the research team exported the building information model from Autodesk® Revit®, added material textures in Autodesk® 3ds Max®, and imported the 3D model into each environment to develop a navigational virtual mock-up. Next, energy performance data was extracted from simulation tools (e.g., EnergyPlus) and incorporated into the virtual mock-up at a component level via information to be displayed in relevant modules. The case study project used in this research was a two-story building, which was built in 1940s. During the time of this study, the building was undergoing an advanced energy retrofit process, where building systems, plug loads and building envelope were analyzed as a whole. The building has been used as a living laboratory to demonstrate energy efficient retrofit methods for Greater Philadelphia region in relation to the Energy Efficient Buildings (EEB) Hub project (EEB Hub, 2012).

IMPLEMENTATION

Important information for integrated design teams to visualize in immersive settings during the evaluation of alternative design options was identified through an iterative process of conducting interviews, developing prototypes and collecting feedback (Frazier et al. 2013). The categories of information that are beneficial for decision-making are listed in the first column of Table 1,which also provides visual forms used for each of the information items and the user interactions that are necessary to trigger the display of the related information.

Table 1. Visual forms and types of information display in both environments

Information to be displayed	Visual forms and types of information display	Examples from the mock-up developed in the commercial VR environment	Examples from the mock-up developed in the game engine environment
Components' thermal properties	*Visual form:* Popup text *Display:* Interaction dependent (when a user hovers mouse over a component)		
Thermal zones	*Visual form:* Image representing zoning map *Display:* Global		
Assumptions for each thermal zone	*Visual form:* Color coded text annotation *Display:* Location dependent (only the thermal zone of the space a user is in)		
Envelope options evaluated for retrofit	*Visual form:* Highlighted buttons *Display:* Global		
Components' R values	*Visual form:* Color coded texture *Display:* Global		
Energy usage intensity	*Visual form:* Animated sliding bar *Display:* Interaction dependent (based on user's selection of retrofit alternatives)		
Initial investment cost			
Annual energy savings			

The selection of these visual display forms was based on an experimentation to explore available functionalities of the two virtual prototyping platforms. These forms were used as-is or tailored to embed and display the necessary information in the virtual mock-up. Screen captures of the visual displays that were used for each of information item listed in Table 1 are illustrated in Figure 1.
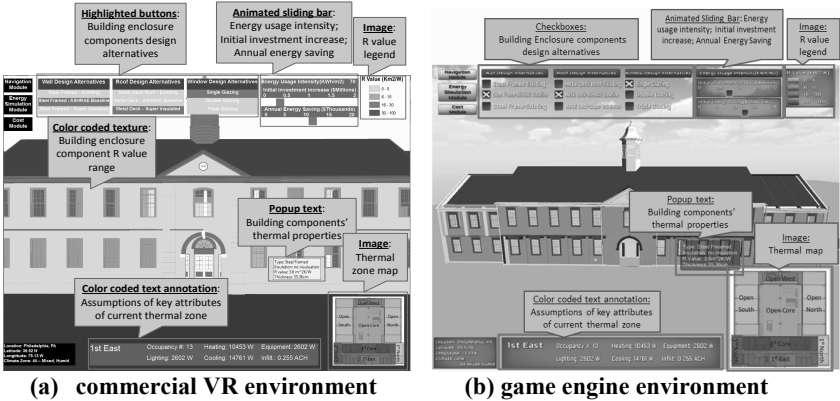
(a) commercial VR environment   (b) game engine environment
Figure 1. Snapshots from the virtual mock-up showing the virtual forms implemented in two environments

Semantic information for several building energy related items was embedded in the virtual mock-ups; however it would result in information overload for users if all of the information were to be displayed on the screen at the same time. The user only needs to see the relevant information at a given time or a space, thus triggers were needed to add functionalities to display necessary information based on the user's interaction with the mock-up. As shown in Table 1, there are three different methods to display information in the mock-up: (1) *Global display*: information will always be displayed on the screen; (2) *User location dependent display*: information will be displayed depending on the user's location in the virtual world; (3) *Interaction dependent display*: information will be displayed if the user performs specific actions in the virtual environment.

LESSONS LEARNED

The virtual mock-up was evaluated in two environments from various aspects and lessons learned about strengths and limitations of these environments were provided below:

1) Simplicity **in converting a BIM to a navigational virtual mock-up:** One of the basic characteristics of a virtual mock-up is enabling user navigation (Kirner et al. 2000). For the purpose of this research, integrated design teams can rapidly develop a virtual mock-up for a 3D facility walkthrough and easily navigate in the virtual mock-up to assess the building architecture for different retrofit alternatives. Both of the environments have built-in functions or modules to enable rapid

development of a navigational virtual mock-up. After importing the 3D model in a compatible file format exported from a 3D modeling software, the user only needs to drag and drop a navigation function/module into the virtual mock-up projects, and navigation through keyboard and mouse will be enabled by default, without additional coding work.

2) Interoperability **with BIM software**: Although converting a BIM to a navigational virtual mock-up is a reasonably established process in both environments, interoperability is an issue. BIM is a semantically-rich information model, while a virtual mock-up generally aims at visualizing only 3D building product geometry. Neither of the two environments can preserve semantic information when converting a BIM to a virtual mock-up. Semantic information stored in a BIM, such as component types, material characteristics and other properties is lost in the file conversion process since all compatible 3D model formats only depict geometric information.

Since sematic information is lost during the conversion of a BIM to a 3D model, there is no parametric modeling feature in the virtual mock-up. In a parametric modeling environment, it is efficient to create a model, for instance, when adding a window to a wall, the wall will be recognized as a host of the window and a void will be automatically created within the wall at the window location. But without this parametric modeling functionality, it is very difficult to modify the existing model. It is possible to modify the virtual mock-up to a limited extent such as rotating and resizing the model components, but this capability of modifying the model is not adequate for representing design revisions. If there is a design revision to the original model, the conversion process must be repeated.

3) Compatibility **in accepting inputs in different formats:** Both of the environments require a 3D model file from modeling software as an input, and both accept widely different file formats. Given the available file formats in the commercial VR environment, the most commonly used file types capable of being generated by BIM or 3D visualization applications are .dwg, .3ds, .dxf, and .VRML. For the game engine based tool, the most commonly used file formats are .fbx, .3ds, .dxf, and .dae.

To include semantic information in the virtual mock-up, the information items in this research were embedded by assigning values to attributes of built-in libraries or user-defined functions. Nonetheless, both of the environments are capable of reading data from text files (.txt) and external databases, which are efficient ways to get semantic information as input from large datasets.

4) Simplicity **in adding user's interaction:** To protect users from being overwhelmed with too much information being displayed at the same time, only the relevant information based on the user's location in, and interactions with the virtual mock-up, are displayed. For example, when a user selects an enclosure option, the color of corresponding building component changes based on the selected alternative's R value. The predefined libraries in the VR software have input and output parameters defined as events that can be linked by drag-and-drop to virtual objects in order to add user's interactions and behaviors. The game engine does not provide a graphical user interface to enable this. By default, all interactions and model behaviors need to be manually scripted in C#, JavaScript, or Boo.

5) Extensibility **through customized functions:** Both of the environments are extensible through implementation of customized functions. For example, the color of text annotation, which shows assumptions of a thermal zone, can be altered based on the current thermal zone that the user is navigating through amidst the virtual environment, and this feature is defined by scripting to get the coordinate of a user's current location and check against the coordinates of the thermal zones. The commercial VR environment supports Jscript, and the game engine supports C#, JavaScript, and Boo to develop these custom extensions.

6) Availability **of built-in libraries to display semantic information**: The identified information, which is critical for evaluating energy retrofit alternatives, should be integrated with the building model. The commercial VR environment has some built-in libraries to be used to develop the Graphical User Interface (GUI), such as tooltip, 2DText and 2DImage, which can be leveraged to embed and display semantic information in the virtual mock-up, as showed in Table 1. The game engine similarly uses 2DText and 2DImage to display information. All GUI elements were manually scripted. However, the GUI scripting in the game engine is quick and straightforward and the GUI scripts can be reused across different projects.

Though both environments have built-in GUI components that can be used to embed and display semantic information, they are generally limited in visualizing semantic information. In this project, manual modifications were needed to the existing libraries to add visual display forms that were not provided. For example, an animated sliding bar was used to show the changes in energy usage intensity, initial investment costs, and annual energy savings as different design options are selected by the users. When a user selects different design options, the markers on the slider will move to match the corresponding number on the scale bars in an animated manner. However, there is no such animated sliding bar in either of the environments. The animated sliding bar used in this virtual mock-up was implemented by building on existing libraries, such as using 2DText to imitate the scale bars and Square to represent the markers. Also, an additional script controls the location of the markers when different alternatives are selected.

7) Repeatability **of the process:** Various visual display forms and interactions were explored in this project, and it is important to understand which part of the process can be repeatedly leveraged in different projects and decisions. In the commercial VR environment, all of the visual display forms and interactions that directly used built-in libraries can be repeatedly utilized in other projects through minor changes to the built-in function attribute values. In addition, the self-defined functions that were developed by scripting in this research work can be copied and re-used in other projects, if the logic is applicable. The game engine performs better in this aspect because it can package all assets used in one project, such as models, scripts, textures, sounds, and GUI elements to be easily imported into other projects.

CONCLUSION

This paper presents lessons learned from experiments performed in two software environments to develop semantically rich immersive virtual mock-ups to support advanced energy retrofit decisions. Processes of developing virtual mock-ups

using a commercially available VR environment and a gaming engine are described, and limitations and advantages of the environments are discussed. Both of the environments have strengths in terms of compatibility with 3D model file formats to accept inputs; simplicity to develop a rapid prototype of navigational virtual mock-up and add user's interaction with built-in functions or modules; and extensibility through scripting customized functions. However, both environments have interoperability challenges with BIM in terms of carrying semantic information and having parametric modeling features. In addition, both of the environments focus on visualizing 3D building product model, but lack information visualization functionalities when it comes to visualize semantic information. The challenges involved in developing these virtual mock-ups in different environments highlights the need for the development of common, modular visualization procedures that can be easily implemented across VR platforms.

REFERENCES

Bowman, D. A., C. North, et al. (2003). "Information-rich virtual environments: theory, tools, and research agenda." Proceedings of the ACM symposium on Virtual reality software and technology.

DoE, U. S. (2008). "Building Energy Data Book." Washington, D.C., Energy Efficiency and Renewable Energy, Buildings Technologies Program, U.S. DoE.

Dunston, Phillip S., and James D. McGlothlin. (2007). "An immersive virtual reality mock-up for design review of hospital patient rooms." CONVR 2007.

Energy Efficient Buildings Hub. (2012). "Navy Yard Building 661: Advanced Energy Retrofit Living Laboratory." Available at: http://www.eebhub.org/media/files/661_final.pdf

Frazier, J., Akinci, B., Ergan, S. (2013). "An approach for capturing requirements of collaborative design teams to facilitate evaluation of energy efficient retrofit design options." Proceedings of Architecture Engineering Institute.

Gopinath, R., and Messner, J. (2004). "Applying immersive virtual facility prototyping in the AEC industry." CONVR 2004.

Heller, J., Heater, M. and Frankel, M. (2011). "Sensitivity Analysis: Comparing the Impact of Design, Operation, and Tenant Behavior on Building Energy Performance." Available at http://newbuildings.org/sensitivity-analysis

Kirner, T. G. and V. F. Martins (2000). "Development of an information visualization tool using virtual reality." Proceedings of the 2000 ACM symposium on Applied computing - Volume 2.

Leicht, Robert M., P. M. P. Moawia Abdelkarim, and John I. Messner. (2010). "Gaining end user involvement through virtual reality mock-ups: a medical facility case study." Proceedings of the CIB W78 2010.

Messner, J. (2006). "Evaluating the use of immersive display media for construction planning." Intelligent Computing in Engineering and Architecture, 484-491.

Maldovan, K. D., Messner, J. I., & Faddoul, M. (2006). "Framework for Reviewing Mock-ups in an Immersive Environment." CONVR 2006.

Whisker, V. E., A. J. Baratta, S. Yerrapathruni, J. I. Messner, et al. (2003). "Using immersive virtual environments to develop and visualize construction schedules for advanced nuclear power plants." Proceedings of ICAPP, vol. 3, pp. 4-7.

Personalized Thermal Comfort Driven Control in HVAC Operated Office Buildings

Farrokh Jazizadeh[1], Ali Ghahramani[1], Burcin Becerik-Gerber[1], Tatiana Kichkaylo[2], Michael Orosz[2]

[1]Sonny Astani Department of Civil and Environmental Engineering
[2]Information Science Institute
University of Southern California, Los Angeles, California, USA, email:
jazizade@usc.edu, aghahram@usc.edu, becerik@usc.edu, tatiana@isi.edu, mdorosz@isi.edu

ABSTRACT

Occupant comfort is a dominant influence on the performance of HVAC operations. Most HVAC system operations rely on industry standards to ensure satisfactory environmental conditions during occupancy. Despite the increasing building energy consumption rates, occupants are not usually satisfied with indoor conditions in commercial buildings. To address this issue, in this paper, a framework for integrating personalized comfort preferences into HVAC control logic is introduced. As part of the framework, a user proxy, a comfort profile learning algorithm, and a building management system (BMS) controller are presented. The performance of the framework in a real building setting has been evaluated. The framework was successful in a small-scale experiment in increasing efficiency by improving user comfort and slight decrease in collective energy consumption.

INTRODUCTION

In the United States, almost half of energy consumption in buildings is consumed by commercial buildings, out of which about 43% is consumed for heating, cooling and ventilation purposes (Department of Energy 2010).Occupant comfort is a dominant influence on HVAC operations. Currently, most HVAC system operations rely on industry standards such as ASHRAE, using the predicted mean vote (PMV) thermal comfort index, to ensure and assess satisfactory environmental conditions during occupancy. The PMV-PPD (predicted percentage dissatisfied) model, which is used as a design criterion to calculate percentage of dissatisfied occupants (e.g., 20% of the building occupants for ASHRAE standards), has been developed based on controlled experiments and introduced in 1970's by Fanger (Fanger 1970). The PMV-PPD is the most frequently used model currently in the industry. However, human comfort perception is a complicated parameter, which is very difficult to be quantified. Fanger's model addressed this complexity by developing a nonlinear model for PMV prediction. The model depends on multiple variables including physical environment variables (e.g. temperature, relative humidity, etc.) and human physiology related variables such as activity level and clothing characteristics. The model was originally developed based on experimental studies and over several decades a number of methods have been developed to facilitate the application of the model. Despite all these developments, the model, used during both the design and

operation of building systems, requires a number of challenging assumptions about the occupant characteristics of the building - such as occupants' metabolic rates, activity rates, and clothing values.

During an operational stage of a building, facility managers conservatively predefine set points in most centrally controlled HVAC operated buildings in the absence of dynamic occupancy information. Application of predefined set points without consideration of dynamic variations could bring about building occupants' dissatisfaction and low efficiency of the building systems (i.e., high percentage of dissatisfied occupants despite high energy expenditures). Various research studies have shown that occupants are not satisfied with predefined thermal comfort set points in HVAC operated buildings (Guo and Zhou 2009). Dissatisfaction with indoor environment could result in reduction in productivity and an unhealthy indoor environment. Accordingly, it is important for building systems to be capable of sensing their occupants' preferences and integrating preferences into operational logic of HVAC systems. In addition to differences in occupants' preferences, individual preferences could vary over time due to seasonal changes, variation in activity level, variation in clothing, etc. A flexible and interactive system is needed to enable a personalized HVAC control system.

To address these issues and propose an approach that accounts for spatiotemporal changes in thermal comfort, this paper presents a framework that integrates complementary comfort preference information to predefined set point model in HVAC operations in office buildings. The core concept of the framework is centered on a user centric sensing approach, in which continuous interaction between a building management system (BMS) and building occupants are required. The framework is part of the authors' human-building interaction vision. The vision argues that accounting for dynamism of occupants in buildings calls for a continuous communication between users and buildings and a mutual behavior change.

BACKGROUND

Improving the performance of HVAC systems in buildings have been the focus of many research studies. In majority of these studies, thermal comfort and building energy consumption have been treated as interconnected parameters and efforts have been made to increase comfort, while conserving energy. Improved understanding of complexities of occupant comfort has led to research in adaptive building control systems. To address the multivariable nature of preferred thermal and air quality conditions, fuzzy logic controllers have been proposed for management of HVAC systems (Calvino et al. 2010; Duan Pei-yong and Hui 2010). Additionally, multi-agent simulations (Dounis and Caraiscos 2009) and neural network computing methods (Atthajariyakul and Leephakpreeda 2005) have been investigated for advanced building controls. Wireless sensor networks have been used for real-time ambient sensing of indoor environments for more informed system controls(Marchiori and Han 2010). Adaptive fuzzy controllers as well as genetic and gradient-based algorithms have been used for multi-objective optimization of occupant comfort and building energy consumption (Kolokotsa et al. 2001; Atthajariyakul and Leephakpreeda 2004). Despite these advances in building control strategies, there has not been a change in comfort models; the PMV index has been

consistently used to assess, maintain, and optimize occupant thermal comfort without validating actual occupant preferences and comfort zones in real building settings.

Addressing this limitation, Guillemin and Morel (2002) has proposed a self-adapting control system that learns specific occupant wishes through user input in the form of set points for thermal and lighting conditions from a keyboard in the test rooms by using an artificial neural network (Guillemin and Morel 2002). Similarly, Daum et al. (2011) proposed a probabilistic measure of thermal comfort to a blind controller using a user interface consisting of too hot/too cold buttons to improve thermal comfort and reduce energy consumption (Daum et al. 2011). Bermejo et al. 2012 applied static fuzzy rules to overall PMV, calculated from temperature and humidity sensor readings in the environment as well as assumptions for activity and clothing levels, as a thermal comfort index. The PMV was used in actuating the room temperature and was updated as occupants interact with thermostat. Through simulations, it was showed that the approach could be used as an online learning method for learning occupants' comfort (Bermejo et al. 2012).

As noted above, the PMV has been used as a thermal comfort index in most of the studies and majority of the proposed methods have been tested in simulation without any attempt to validate proposed approaches in real buildings. Moreover, constant values have been assumed for clothing and activity level values. Accordingly, in this paper, authors proposed an approach that enables a user centered HVAC control in existing buildings and addresses the noted issues by using the industry standard building systems' design as a baseline and integrated user input as the complementary information to improve HVAC system efficiency.

PERSONALIZED THERMAL COMFORT BASED HVAC CONTROL

The proposed approach is based on the fact that existing building systems (designed based on the industry standards) provide a baseline indoor environment for thermal comfort; however, the dynamism with regards to occupants (e.g., change in activity, metabolic rates, etc.) are not taken into account during the operational stage. To address this issue, the author's proposed approach uses participatory sensing. In participatory sensing, users are the sensing nodes, which provide the context dependent spatiotemporal information of the actual events in buildings.
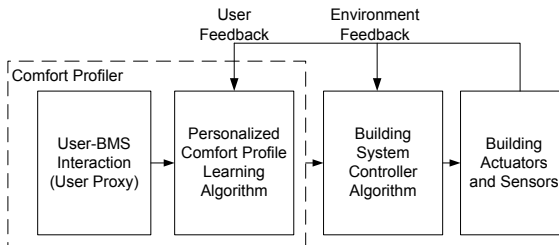
Figure 1. Personalized HVAC Control Framework

In fact, the proposed approach enables users to communicate their preferences (which stems from environmental, physiological, and psychological factors) to BMS. Accordingly, the approach requires using a proxy, which represents building users.

An actuating mechanism is also designed to enable BMS reaction to users' preferences (Figure 1).

User-BMS Communication Proxy. In the proposed framework, users are able to communicate their thermal comfort votes to BMS using ubiquitous computing devices. Smartphones, tablets, laptop or desktop computers could be used as platforms (for user proxy) for context dependent thermal comfort information. Accordingly, an efficient design of the interface is required to enable user-BMS communication. Proxy design (the design of the interface features) was carried out to enable users to provide feedback that is consistent and represents their true perceptions of the environment and their satisfaction with the indoor environmental conditions. Consistency in this context was defined as similarity of the votes under similar indoor ambient conditions. Accordingly, proxy design has been carried out through a series of task execution surveys, think-aloud cognitive design evaluations, and interviews. Details of the interface design are not part of the scope of this paper.
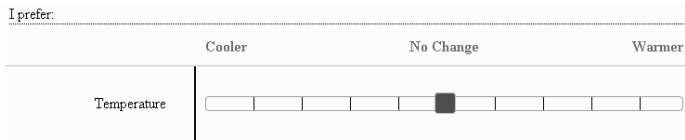
Figure 2. Thermal vote preference slider on the proxy interface

Interface design studies showed that the most effective interface feature for capturing users' consistent thermal vote is a preference scale (compared to the ASHARE thermal comfort sensation perception scale) with snapping. The preference scale asks for user's reaction to current indoor temperatures. Correlation analyses on the field captured data showed that temperature is the most important factor in driving users' vote. Based on these findings, a proxy has been designed for different platforms such as iOS and Android based smartphones and personal computers.

Learning Users' Comfort Profiles. Using the interface (figure 2), users are able to communicate their individual preferences to BMS. HVAC system reacts to users' vote by adjusting indoor conditions to meet user's thermal comfort requirements. As users interact with the system to adjust the temperature, they also provide information to BMS, which is used to learn their comfort profiles. Accordingly, another component of the proposed framework is a network of wireless sensors that capture ambient conditions in the rooms of interest. In this study, the sensor system consists of an Arduino Black Widow stand-alone single-board microcontroller computer with integrated support for 802.11 WiFi communications and MaxDetect, RHT03 temperature/humidity sensor. The measurement accuracy is ±0.2°C and the resolution (sensitivity) is 0.1°C.

Each user's thermal comfort feedback and ambient conditions for which the feedback is provided are captured. The interface design helps the BMS to capture the preference (how much change in temperature) and perception (an index which represents discomfort) at the same time. In order for users to interact with BMS, an initial linear relationship between slider values (a number between -50 and 50) and requested temperature change are used. However, to avoid excessive trial and error for finding the optimum temperature range and also customize the slider for each

user, the proposed framework uses a learning algorithm to determine comfort profile of each user. Comfort profile, in this context, is the relationship between users' comfort indices and ambient temperature. As noted, comfort perception is a quantity that is affected by numerous factors and therefore users report similar thermal comfort feedback inputs under different indoor conditions. Thus, different thermal comfort indices are associated with a range of ambient temperatures. Accordingly, to learn users' comfort profiles a fuzzy rule based data mining approach, namely the Wang-Mendel (WM) model (Li-Xin Wang 2003), was used in this study. Fuzzy logic is used to simulate human subjective thinking. In this study, instead of introducing the rules by human experts, rules are learned from data provided by human subjects (users); thus the rules represent each individual's comfort profile. The mathematical representation of the learning problem is defined as a set of data points $\{(tp_1, t_1), (tp_2, t_2), \ldots, (tp_N, t_N)\}$ where tp_i is the thermal perception index as input and t_i is the associated ambient temperature for each individual. The objective is to determine a mapping ($f: tp \rightarrow t$) between thermal perception indices and the ambient temperature as fuzzy rules, which categorize temperature ranges in the form of fuzzy sets and use the generated fuzzy rules for predicting. Accordingly, while user interacts with HVAC system to adjust the preferred temperature, his/her comfort profile is learned and a customized sensation scale (the relationship between preferred temperature and the slider value) for each individual is developed. In addition to the mapping between comfort indices and temperature, thermal comfort ranges for each individual is obtained.

BMS Controller. One of the main objectives of the framework is to improve building systems efficiency in existing buildings. Accordingly, the framework is designed to require no additional hardware to the HVAC system and minimum intrusion to the building. However, to improve the thermal comfort experience, room level temperature (compared to the zone level temperature in native HVAC system) is taken into account. In order to react to users' preferences, a proportional controller is used. The proportional controller reacts to the feedback from the environment to keep temperatures of occupied rooms as close as possible to the preferred temperatures as reported by the comfort model. The controller aims to minimize the sum of deviations from preferred temperatures, which serves as the definition of error:

$$error = \frac{\sum_{i=1}^{N_r}(T_r^i) - (T_p^i)}{N_r}$$

where N_r is the number of rooms in a thermal zone, (T_p^i) is the preferred temperature and (T_r^i) is the measured temperature in each room of a thermal zone. The dynamic control mechanism inherently accounts for factors such as heat loads in rooms, differences in solar heat gain, and temperature reading offsets due to the location of different sensors by dynamically driving the system control to minimize the error, as opposed to trying to compute a "perfect" set point.

EXPERIMENTAL STUDY

Objective. The objective of the experiment is to evaluate the performance of the framework in a real building setting through the assessment of thermal comfort and energy consumption consequences of the framework. Energy consumption-comfort improvement trade off was explored.

Experiment Test Bed. The proposed framework has been deployed in an office building on a university campus in the Southern California area. The framework has been activated for two zones of the building. Each zone includes two single occupancy office rooms and each room was equipped with ambient temperature sensors. Users in each room were provided with the thermal comfort app (as a proxy). The results are presented for one user from each zone (one room from each zone).

Comfort Profile Development. Experiments have been started in mid October 2012 and continued for three weeks. The first two weeks have been considered as the training period, in which users interacted with the system to adjust the temperature and provide training data to BMS. The comfort profiles of the users in the two- targeted rooms are presented in Figure 3. The evolution of the comfort profiles over the time has been shown in this figure. The number of fuzzy sets is a model parameter, which determines the accuracy of the fuzzy rule extraction method. Results of the sensitivity analysis showed that on average five fuzzy sets result in a global optimum error. However, depending on individual users the number of fuzzy sets could vary between 3 and 7. As it is shown in Figure 3, by increasing the number of data points the resolution of the comfort profile can be increased.

| Room 1 - comfort range learning: upper) comfort ranges; lower) sensation scale | Room 2 - comfort range learning: upper) comfort ranges; lower) sensation scale |
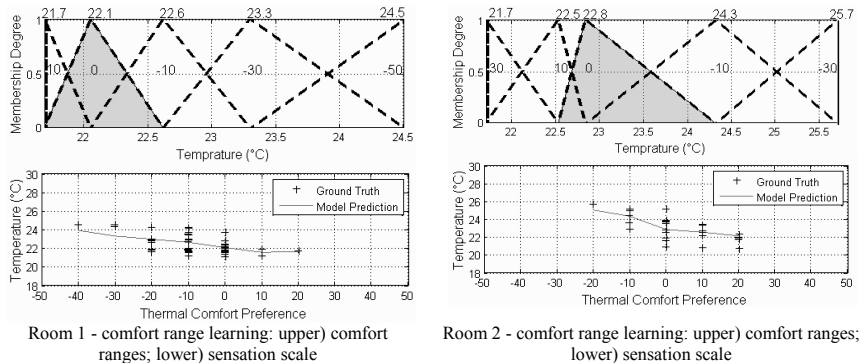
Figure 3. Learned personalized comfort profiles

Upon obtaining the comfort profiles of users during the first two weeks, the BMS controller started using personalized comfort profiles by setting the default preferred temperature of each room to the equivalent to users' preferred temperature. Users' preferred temperature is the temperature equivalent to the center of the comfort profile (associated with zero votes). The requested temperature changes were also calculated based on the personalized sensation scale.

Comfort and Energy Tradeoff. Performance of the framework in providing comfort was evaluated through surveys with the users. To quantify the satisfaction of users with the proposed framework, users were asked to rank their satisfaction of the indoor environment from 1 to 10 (10 being extremely satisfied) before and after the deployment of the framework. The reported ratings, before and after of the one-to-one exit interviews were done and the users expressed their satisfaction with the deployed framework by mentioning that the preferred temperatures are compatible with their expectations and they prefer the system to keep operating.

In order to evaluate the effect of the framework on energy consumption in the targeted zones, a benchmarking test for comparison was conducted. Energy consumption could be calculated for different components such as airflow for maintaining indoor environmental conditions and energy for cooling and heating the air. In this study, variation in the airflow has been used as a metric for comparing energy consumption. Energy consumption in the target zones for two periods (before and after the framework) was compared. In selecting the periods for comparison, having similar outside temperature conditions was a criterion. Outside temperature ranges for the comparison period were 17.4 to 27.2 °C and 16.6 to 26.4 °C for period 1 and period 2, respectively. Average airflow for one week during the tests was measured and compared with the similar period using the control data. Results are presented in Table 1. Results show that in some cases energy savings could be achieved and in some cases energy consumption increases depending on user preferences. However, part of the increase in energy consumption is due to the oscillations in airflow values. This is related to the fact that the proportional controller needs to be tuned. If the model is tuned, there is potential for saving energy. In general, when the HVAC system is in the cooling mode, energy savings could be achieved when users are asking for a warmer environment. According to authors' observations, majority of the dissatisfied users in building prefer warmer indoor conditions, which shows the potential for saving energy. Energy-comfort tradeoff will be explored with a large-scale experiment, as part of authors' future work. The proposed framework has the potential to increase efficiency (improving comfort with no increase in energy consumption) of building systems depending on the HVAC operational mode and users' preferences.

Table 1. Comparison of air-flow in targeted zones

Zone	Before Framework (c^3/m)	After Framework (c^3/m)	Change (%)
Zone 1	9.64	11	+14%
Zone 2	9.57	7.93	-17%

CONCLUSIONS

A framework for integrating user preferences in the control loop of the HVAC system in office buildings was presented. The main objective of the framework is to account for human factors that affect thermal comfort in buildings such as activity levels and clothing values. Participatory sensing approach using ubiquitous computing devices, a fuzzy rule extraction algorithm, and a proportional controller were used as components of the framework. Results of the experiments in two zones of an office building showed the efficacy of the framework in learning personalized comfort profiles, improving comfort, and maintaining energy consumption. More comprehensive benchmarking of energy consequences of the framework, implementation of the framework at building level, improvements to the controller performance, and formulation of an optimization problem for air flow are among authors' future research tasks.

ACKNOWLEDGMENTS

Department of Energy or the National Science Foundation.

REFERENCES

Atthajariyakul, S., and Leephakpreeda, T. (2005). "Neural computing thermal comfort index for HVAC systems." *Energy Conversion and Management,* 46(15-16), 2553-65.

Atthajariyakul, S., and Leephakpreeda, T. (2004). "Real-time determination of optimal indoor-air condition for thermal comfort, air quality and efficient energy usage." Elsevier Ltd, 720-733.

Bermejo, P., Redondo, L., Ossa, D. L., Rodriguez, D., Flores, J., Urea, C., Gamez, J. A., and Puerta, J. M. (2012). "Design and simulation of a thermal comfort adaptive system based on fuzzy logic and on-line learning." *Energy Build.,* 49 367-379.

Calvino, F., La Gennusa, M., Morale, M., Rizzo, G., and Scaccianoce, G. (2010). "Comparing different control strategies for indoor thermal comfort aimed at the evaluation of the energy cost of quality of building." *Appl.Therm.Eng.,* 30(16), 2386-2395.

Daum, D., Haldi, F., and Morel, N. (2011). "A personalized measure of thermal comfort for building controls." *Build.Environ.,* 46(1), 3-11.

Department of Energy, . (2010). "Buildings Energy Data Book available at http://buildingsdatabook.eren.doe.gov/ChapterIntro1.aspx." *Last Accessed June 2011,* .

Dounis, A. I., and Caraiscos, C. (2009). "Advanced control systems engineering for energy and comfort management in a building environment-A review." *Renewable and Sustainable Energy Reviews,* 13(6-7), 1246-1261.

Duan Pei-yong, and Hui, L. (2010). "A novel data-based control strategy of dynamic thermal comfort for inhabited environment." *2010 8th World Congress on Intelligent Control and Automation (WCICA 2010),* IEEE, Piscataway, NJ, USA, 4865-9.

Fanger, P. O. (1970). "Thermal comfort : analysis and applications in environmental engineering." .

Guillemin, A., and Morel, N. (2002). "Experimental results of a self-adaptive integrated control system in buildings: a pilot study." *Solar Energy,* 72(5), 397-403.

Guo, W., and Zhou, M. (2009). "Technologies toward thermal comfort-based and energy-efficient HVAC systems: a review." *SMC 2009,* IEEE, Piscataway, NJ, USA, 3883-8.

Kolokotsa, D., Tsiavos, D., Stavrakakis, G. S., Kalaitzakis, K., and Antonidakis, E. (2001). "Advanced fuzzy logic controllers design and evaluation for buildings' occupants thermal-visual comfort and indoor air quality satisfaction." *Energy Build.,* 33(6), 531-43.

Li-Xin Wang. (2003). "The WM method completed: a flexible fuzzy system approach to data mining." *IEEE Trans.Fuzzy Syst.,* 11(6), 768-82.

Marchiori, A., and Han, Q. (2010). "Distributed wireless control for building energy management." *2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, BuildSys'10, November 2, 2010 - November 2,* Association for Computing Machinery, Zurich, Switzerland, 37-42.

Integration of Change and Knowledge Management Processes in Energy Efficient Retrofit Projects

Fangxiao Liu[1], Abdou Karim Jallow[2*], Chimay J. Anumba[3] and John I. Messner[4]

[1,2,3,4] Department of Architectural Engineering, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802, USA; Email{fzl116; akj10; anumba; jmessner}@psu.edu

ABSTRACT

Managing change and knowledge is one of the key factors that contribute to a successful energy efficient retrofit project. Due to the constraints of information, time, space and working environment, a retrofit project is more complex than expected. The significance of managing changes and dependencies is that if they are poorly or inadequately managed, changes initiated and implemented during the design, construction and operations of buildings could potentially impact negatively on energy efficiency goals. Similarly, knowledge assets are often disconnected from project teams. Consequently, lack of access to the right knowledge by project teams could result in the design and retrofitting of energy efficient buildings which do not meet the owner's goals. Furthermore, there is potential for lessons to be learned during the process of managing and implementing changes. However, this kind of information is usually not properly captured, validated and stored within a knowledge management system (KMS). To do this more effectively, the change and knowledge processes need to be integrated at the points of interactions. This paper presents an approach to integrate change and knowledge management (CKM) processes. It highlights the importance of the CKM practices in energy efficient retrofit projects, and describes how lessons learned from change events can be captured and stored as knowledge in a KMS. Furthermore, it explains how a KMS can be expanded into a change and knowledge management system (CKMS) to support the new approach. In the concluding part of the paper, the key benefits of the proposed approach are highlighted and the approach to be taken towards the practical implementation of a CKMS is outlined.

INTRODUCTION

A construction project is unstable and complex, and involves various participants (Jia, 2010). A large proportion of the problems on construction projects emanate from the lack of appropriate mechanisms to establish and track the dependencies between project elements, which makes it very difficult, and in some cases, impossible to anticipate and manage the changes that occur in the course of a project (Jallow, 2011). This can also have potentially debilitating impacts on a project thus impairing the likelihood of successful project outcomes. Similarly, the inadequacies in systems and approaches to managing knowledge and the lessons learned on projects often result in considerable rework, repeated mistakes, 'reinventing the wheel', unnecessary expenses and other inefficiencies. Given that changes on projects often serve as a learning opportunity, there is a need for systems that can simultaneously manage change and knowledge. Project teams are usually aware of the importance and necessity of knowledge generation and utilization, as

well as managing changes. However, they generally allocate a lower priority to these activities than to primary project goals.

A retrofit project is "the modification or conversion (not a complete replacement) of an existing process, facility or structure. Such modification may involve additions, deletions, rearrangements or not-in-kind replacement of one or more parts of the facility. Changes may alter the kind, quantity, cost or quality of the products or services being produced by the facility." (Sanvido and Riggs, 1993, pp.ii) The significance of managing changes and dependencies in the context of energy efficient retrofit projects is that if they are poorly or inadequately managed, changes initiated and implemented during the design, construction and operations of buildings could potentially impact negatively on energy efficiency goals. A large number of modeling simulations need to be done in order to get the most optimized design which will meet energy efficiency goals (Heo et al., 2012). Similarly, appropriate knowledge is also needed to determine the most optimized design. Without knowledge, people will make decisions based on assumptions and experience (Heo et al., 2012). Lack of access to the right knowledge by project teams could result in the design and retrofitting of buildings which does not meet the owner's goals for energy efficiency.

RELATED WORK

Change and Dependency Management (CM)

A change refers to "an alteration or a modification to pre-existing conditions, assumptions or requirements" in construction work (Hao et al., 2008, pp.4). Problems caused by changes are typically related to schedule, efficiency, cost, and quality (Sun et al., 2004; Zou and Lee, 2008). Dependency is the interactions among different elements, and a change in one element must coordinate with others to achieve the functionality of a system (Jallow, 2011). Change management (CM), according to Voropajev (1998), is "an integral process related to all project internal and external factors, influencing project changes; to possible change forecast; to identification of already occurred changes; to planning preventive impacts; to coordination of changes across the entire project" (pp.17). Dependency management is aimed to "provide a procedure and associated guidelines to facilitate the management of project dependencies" (COGTA, 2006, pp.1). Motawa et al. (2007) indicated that identification of change sources, evaluation of change effects and dependency management are critical activities of dependency management, and Leyland et al. (2009) emphasized this by stating that dependency management is used to analyze, control and mitigate various factors in construction processes. Sun et al. (2004) also highlighted that adequate management of changes is necessary to avoid (or mitigate) the consequences of direct and indirect impacts.

Knowledge Management (KM)

According to Stewart (1998), knowledge has become the most important factor and resource for organizations today. Knowledge is dynamic justified true belief and is created by individual as well as organization interactions in the society (Nonaka et al., 2000).

Knowledge management (KM) helps in the planning and organizing companies' knowledge assets to improve performance and production efficiency (King, 2007). Generally, knowledge management identifies and analyzes necessary available knowledge, therefore generating actions to fulfill organizational/individual goals by developing knowledge assets (Sensky, 2002). It is designed for sharing an organization's expertise to make knowledge explicit, and managing information with the help of information technologies (Ribino et al., 2009). Knowledge management plays an important role to provide organizations with competitive advantage (Elgobbi, 2010). According to the survey conducted by KPMG (2003), 80% of the participating organizations considered knowledge as strategic assets, and many companies implied that 6% of their annual revenue was lost due to the failure to implement knowledge management. In energy efficient retrofit projects, information is an apparent limitation. The existing site conditions may be uncertain, the project scale may be ill-defined, and design drawings may be out of date (Sanvido and Riggs, 1993). Therefore, knowledge management is important for a retrofit project.

There are a number of knowledge management systems that have been developed (Anumba and Pulsifer, 2010), such as the CAPRI.NET system, CLEVER_KMTM, and IMPaKT, etc. The CAPRI.NET system is web-based database for project knowledge management (Udeaja et al., 2008), and it is developed in a research project of knowledge live capture (Tan et al., 2010). It can store information in the format of project knowledge file (PKF), and team members can share their project knowledge through the system (Tan et al., 2010).

INTEGRATION OF CM AND KM

The Need for an Integrated Approach

It is fundamental and extremely relevant that change and knowledge management processes are integrated to support the integrative energy efficient design and development. Multi-scale and multi-discipline interactions occur during design and development with several decisions made. Often, these decisions result in some changes to project variables. As a result of dependencies between project elements, this also creates the phenomenon of change propagation by which a change to a project scope or design element requires additional changes throughout the building. Furthermore, there is also potential for lessons to be learned when performing project activities and during the process of managing and implementing changes. However, knowledge learned from this process is not, if at all, adequately captured and put in the knowledge management system. Consequently, it is important that lessons learned are captured, validated and stored for use in later stages and future projects. This creates a bi-directional relationship between change and knowledge management processes. This means change management processes are application areas of knowledge and within these processes, knowledge is generated through capture and validation processes of lessons learned. The integration of change and knowledge management will significantly contribute to the project variables such as time, budget, value and benefits. This will make provision for the dynamic management of change and knowledge, which are usually static and isolated from the workflows, and propels the ability and agility of project teams and building

managers to respond appropriately to changes (however initiated) and new knowledge. This agility includes being able to approach and manage problems in a proactive manner.

Figure 1 shows the integrated change and knowledge management (CKM) approach. When a change happens, all related lessons learned and sources will be collected by the knowledge management system, which will be stored in the knowledge repository as new knowledge after validation. On the other hand, knowledge captured during the construction process can be used to predict potential changes in the project, thus taking measures in advance. During the change resolution stage, the knowledge management system can act as a platform to call for the most suitable person to help solve problems caused by the change, and lessons learned can be captured back to the knowledge management system. Finally, all change/dependency related knowledge will be stored properly and shared among different teams.
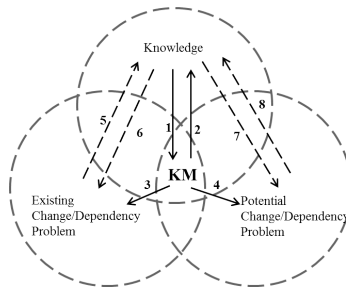
(1-Capture, 2-Share, 3-Resolution, 4-Predict, 5&8-Lessons learned, 6&7-Sources)
Figure 1. Integration of CM and KM

Implementation of the Integrated Approach

High level CM and KM process maps have been developed by adapting previous work of Motawa et al. (2006), Tan et al. (2010) and the EPSRC Industrial Report (Sun et al., 2004) using Business Process Model and Notation (BPMN) method, which provides a notation approach to standardize business process model design and process implementation (OMG, 2011). The CM process map has 17 activities in 4 stages: Start Up, Identify & Evaluate, Approval, and Implement & Review. Similarly, the KM process map has 8 activities in 4 stages: Identification, Capture & Store, Share & Use/Reuse, and Maintenance.

The integration of change and knowledge management processes was implemented by using BPMN *signal intermediate events* (see Figure 3 & 4). According to BPMN method, *signal intermediate events*, which include *'send'* and *'receive'* (see Figure 2), are used for sending or receiving general communication within and across processes during the flow of a process. The *signal send intermediate event* is used to send signals, and the *signal receive intermediate event* receives signals, which trigger the process to continue (OMG, 2011). For example, when changes are requested, approved and implemented, lessons learned will be captured and the *signal send intermediate event* (i.e., 'Change-driven knowledge', as

highlighted in Figure 3) on the change management process will be triggered to communicate the captured knowledge as a signal to the knowledge management process. The *signal receive intermediate event* (i.e., 'Change-driven knowledge', as highlighted in Figure 4) will receive the broadcasted signal and trigger the knowledge management process. In this way, the lessons learned from changes can be transferred and stored adequately in the knowledge management system following their validation. Without this integrative approach, it is very likely that the lessons learned from changes would not be captured and made available in the KMS for future utilization, thus causing loss of knowledge. The same approach can be applied to retrieve knowledge from the KMS to utilize during change approval processes.

Signal Send
Intermediate Event

Signal Receive
Intermediate Event

Figure 2. Signal Intermediate Events
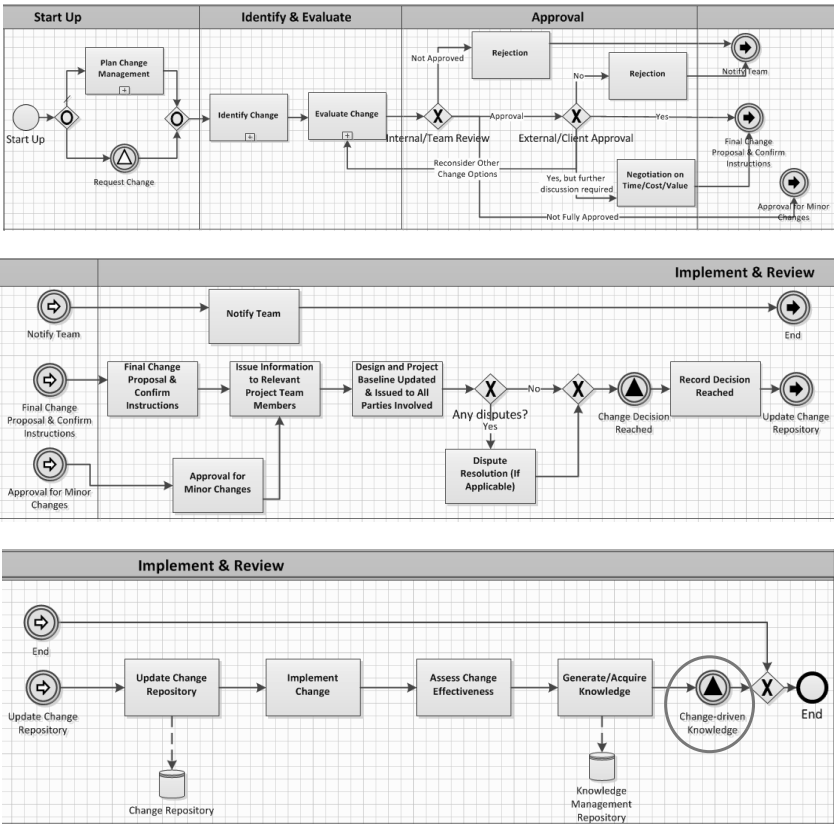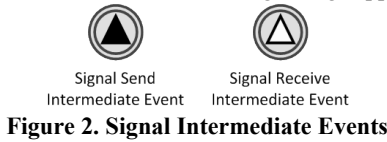
Figure 3. CM Process Model with Signals
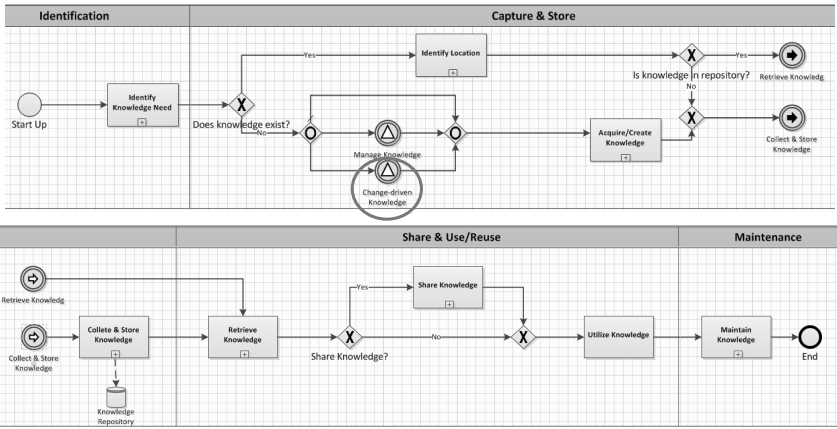
Figure 4. KM Process Model with Signals

Proposed System and Benefits

The paper proposes a Change and Knowledge Management System (CKMS) by developing a change management module (CMM) within the Knowledge Management System (KMS). The purpose of this module is to serve as a repository of change information embedded in the KMS from which potential knowledge can be derived. In this way, it is possible to avoid the loss of potential knowledge during change management processes. The research reported in this paper has identified a knowledge management system called CAPRI.NET to be used for the project. This is intended for the 'live' capture and reuse of project knowledge and it is planned to extend this through the development of the change management module, which enables project teams both to effectively address any changes that occur on the project and to capture any associated lessons learned from those changes.

The integrated CKM approach is being adopted and demonstrated within the integrated building lifecycle processes in the $129m Energy Efficient Buildings Hub (EEBHub) project from the Department of Energy (DOE) and the Commonwealth of Pennsylvania. The following include benefits associated with integration of change and knowledge management processes.

- Facilitate information flow and exchange between teams involved in change and knowledge management.
- Establish a systematic, methodical and standardized way for the management of changes and knowledge for value creation.
- Seamlessly capture lessons learned during the change management process, and subsequently achieve them in a KMS.
- Help avoid or mitigate against loss of potential knowledge.
- Utilize knowledge for key decision-making during change approval processes.

CONCLUSIONS

This paper presented an integrated CKM approach to combine CM and KM processes. It also proposes an integrated system (CKMS) based on the integrated

approach by embedding a CM module into a KMS, CAPRI.NET. Energy efficient retrofit projects can benefit significantly from the implementation of the CKM approach and CKMS. This can offer the potential to considerably improve the simulation and analysis process, and enhance the industry's performance by adequately managing the change as well as the knowledge processes. The limitation of the paper is that the integrated approach presented is theoretical and no practical implementation has been done. Since the concept is based on validated approaches, it is anticipated that when implemented, substantial improvement will be achieved towards energy efficient retrofit projects. In the future, the CKMS will be developed, operationalized and validated in retrofit projects. Industry evaluation of the integrated approach will also be conducted to ascertain its applicability. It is also recommended that the integrated CKM approach be embedded within the whole lifecycle processes of a project.

ACKNOWLEDGEMENT

REFERENCES

Anumba, C.J. & Pulsifer, D. (2010). "Knowledge Management Systems for Construction." In: Construction Research Congress 2010.

COGTA. (2006). "Dependency Management."
<http://www.cogta.gov.za/subwebsites/mig/toolkit/TOOLBOX/PM/Dependency%20Management.pdf> (Accessed November 2012).

Elgobbi, E.M. (2010). "Exploring the perceptions of Knowledge Management and Knowledge Management Systems: A case study on the Libyan oil industry." *Advanced Management Science (ICAMS), 2010 IEEE International Conference* on, pp.445.

Hao, Q., Shen, W., & Neelamkavil, J. (2008). "Managing Changes in Construction." NRC Report –RR-258.
<http://www.nrccnrc.gc.ca/obj/irc/doc/pubs/rr/rr258/rr258.pdf> (Accessed March 2012).

Heo, Y., Choudhary, R. & Augenbroe, G. A. (2012). "Calibration of building energy models for retrofit analysis under uncertainty." *Energy and Buildings*, Vol. 47, pp.550-560.

Jallow, A. K. (2011). "Integrated Lifecycle Requirements Information Management (eRIM) in Construction Projects." PhD Thesis. February 2011, Loughborough University (Unpublished).

Jia, W. (2010). "Application of project management techniques in construction management." In: 2010 *IEEE International Conference on Advanced Management Science (ICAMS)*. Presented at the 2010 IEEE International Conference on Advanced Management Science (ICAMS), IEEE, pp. 516–520.

King, W.R. (2007). "Knowledge Management: A Systems Perspective". International Journal of Business and Systems Research, 1(1), pp.5-28.

KPMG (2003). "Insights from KPMG's European Knowledge Management Survey 2002/2003". UK: KPMG's European.

Leyland, M., Hunter & D., Dietrich, J. (2009). "Integrating Change Management into Clinical Health Information Technology Project Practice". In: World Congress on Privacy, Security, Trust and the Management of e-Business, 2009. Presented at the World Congress on Privacy, Security, Trust and the Management of e-Business, 2009. CONGRESS '09, IEEE, pp. 89–99.

Motawa, I.A., Anumba, C.J. & El-Hamalawi, A. (2006). "A fuzzy system for evaluating the risk of change in construction projects." Advances in Engineering Software 37, pp. 583–591.

Motawa, I.A., Anumba, C.J., Lee, S. & Peña-Mora, F. (2007). "An integrated system for change management in construction." Automation in Construction, Vol. 16, pp. 368–377.

Nonaka, I., Toyama, R. & Konno, N. (2000). "SECI, Ba and Leadership: a Unified Model of Dynamic Knowledge Creation." Long Range Planning, 33(1), pp.5-34.

OMG. (2011). "Business Process Model and Notation (BPMN)." <http://www.omg.org/spec/BPMN/2.0/> (Accessed October 2012).

Ribino, P., Oliveri, A., Re, G. L. & Gaglio, S. (2009). "A Knowledge Management System Based on Ontologies." In: International Conference on New Trends in Information and Service Science, 2009. International Conference on New Trends in Information and Service Science, 2009. IEEE, pp. 1025-1033.

Sanvido, V. E. & Riggs, L. S. (1993). "Managing Successful Retrofit Projects." Cost Engineering, Vol. 35, pp. 25–31.

Sensky, T. (2002). "Knowledge management. Advances in Psychiatric Treatment." 8, pp. 387-396.

Stewart, T. & Ruckdeschel, C. (1998). "Intellectual capital: The new wealth of organizations." Performance Improvement, 37(7), pp.56-59.

Sun, M., Sexton, M., Aouad, G., Fleming, A., Senaratne, S., Anumba, C., Chung, P., El-Hamalawi, A., Motawa, I. & Yeoh, M. L. (2004). "Managing Changes in Construction Projects." EPSRC Industrial Report, University of the West of England: MCD Research Project. <http://www.bne.uwe.ac.uk/cprc/publications/mcd.pdf> (Accessed October 2012).

Tan, H. C., Anumba, C. J., Carrillo, P. M., Bouchlaghem, D., Kamara, J. & Udeaja, C. (2010). "Capture and Reuse of Project Knowledge in Construction". John Wiley and Sons.

Udeaja, C.E., Kamara, J. M., Carrillo, P. M., Anumba, C. J., Bouchlaghem, N. D. & Tan, H. C.(2008). "A web-based prototype for live capture and reuse of construction project knowledge." Automation in Construction, 17(7), pp.839-851.

Voropajev, V. (1998). "Change Management—A Key Integrative Function of PM in Transition Economies." International Journal of Project Management, Vol. 16, pp.15–19.

Zou, Y. & Lee, S. (2008). "The impacts of change management practices on project change cost performance." Construction Management and Economics 26, pp.387–393.

Process-based Information Exchanges Mapping for Energy Efficient Retrofit Projects

Abdou Karim Jallow[1], Sanghoon Lee[2], Fadi Castronovo[3], Ying Zhang[4], Sreelatha Chunduri[5], and John I. Messner[6]

[1, 2, 3, 5, 6]Department of Architectural Engineering, The Pennsylvania State University, University Park, PA16802; email {akj10, sul37, fxc175, szc192, jmessner}@psu.edu
[4]Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA16802; email: yuz143@psu.edu

ABSTRACT

Access and exchange of building information in a timely and standard manner between construction project teams and applications during design and construction, plays a major role in assuring efficient building energy simulation during retrofitting. Yet, in spite of several efforts in major technological developments in information management, building data and information remains highly fragmented and frequently inaccessible to project teams. This paper presents a table for categorization of information exchanges items and illustrates their dependencies between internal process activities, and between activities of different processes. The paper uses the HVAC design process as a baseline to map-out the information *input-and-output* relationships of various activities and shows how an information item produced by an activity is consumed by another, thus creating information exchanges dependency. The table can serve as a reference point for project teams and for software developers to understand interoperability requirements of process activities and applications. It will also be beneficial to analyze impact of changes as change to an information exchanges item can impact on other activities where it is required as an input. The paper concludes by formulating a roadmap to utilize the table in order to support development of Information Delivery Manual (IDM) and other information exchanges initiatives.

INTRODUCTION

It is generally believed that exchange of relevant information is extensive and one of the main challenges in design, construction and operations of buildings (Hjelseth, 2010). This is the case for both new and retrofit projects. However, it is more demanding in energy efficient retrofit projects where varying number of simulations and analyses are carried out requiring multiple exchange of information between heterogeneous applications. During this process, several activities are carried out for the design of each building element/system and sequencing these activities is influenced by the information dependency (Maheswari and Varghese, 2005). The activities often interact internally and externally with other building elements/systems design activities. Inadequate exchange of information could potentially impact the analyses adversely.

The need to integrate processes and applications seamlessly for building energy simulations during design is colossal. The various domain specific analyses, e.g., HVAC, structural and lighting can very effectively make accurate analysis and energy efficiency related decisions based on effective exchange of information. A major challenge in seamless exchange of information is the information incompatibility problem as different models and formats are used by different actors, thus perpetuating the fragmentation of information flow (Kriphal and Grilo, 2012). Seamless exchange of information of different formats between the specific domains, and how to specify the exchange requirements for the various simulation activities remain challenging. This (i.e., seamless

exchange of information) is a driving force for application integration, a process which requires the processing of relevant and necessary information between the domain specific processes and applications. This is necessary in particular when adopting a BIM-based (Building Information Model) information exchanges. As highlighted by Hjelseth (2010), BIM standards should contain definitions of mandatory and optional information related to the different roles and phases in the building process through its life cycle. However, identifying and mapping out the specific information is also challenging. Categorized documentation of the different types of information consumed, produced, and exchanged by the various process activities is not available. No known studies have been found to actually identify and tabulate the specific information exchanges item types and dependencies of process activities in this manner. Thus, the aim of this paper is to develop a tool (table) that can be utilized to identify and categorize the various types of design information exchanges that are utilized and produced during lifecycle process execution. To demonstrate its utilization, the paper presents an information exchange table of HVAC design process. The scope of the paper is limited to the information exchanges at the design phase only. However, the table can be used for all other phases too. The remaining parts of the paper provide background study, research methodology, information exchanges mapping approach and a proposed tool (table). The last section summarizes the paper, and defines future work in this area.

RELATED WORK

Process and information exchanges

Construction processes define the activities performed to deliver a product or a service, and to increase process understanding is to consider their parts and relationships; one way is to analyze the information flow (Zhao, Lv and You, 2008). To establish effective collaboration between multidisciplinary project teams, the organizations involved need to capture and share product development information and knowledge in the whole lifecycle processes (Geryville et al., 2005). In particular, during energy efficient retrofit design and construction, multi-scale energy modeling and simulations are conducted to define an energy efficient solution which meets energy goals. This involves the use of information and models of varying scale and sources. Information and communication technologies (ICT) have widespread applications in the construction industry over the past years. Amongst other factors, the widespread application is driven by the need to bridge the fragmentation which characterizes the industry and to facilitate effective management of information (Brewer and Gajendran, 2009). Specific systems or applications, often heterogeneous are required for individual activities, and this creates integration and interoperability difficulties between the applications. To develop the specific applications, it is crucial to consider the processes, activities and information that will be processed. Han and Froese (2009) suggest that a task/transaction/integration approach be taken where consideration is given to each project task, each transaction involving information flows between tasks, and significant integration issues to arrive at a list of information systems requirements. In their work on process based information architecture (PBIA), Sanvido et al. (1995) identified and classified the information required to manage, plan, design, construct and operate a facility. This was accomplished by analyzing and classifying the information flow between the lifecycle processes, however, PBIA did not define the individual information exchanges items for each process activity according to building elements and design phases. Recent efforts in facilitating effective BIM-based information exchange show the development of the Information Delivery Manual (IDM) and Model View Definitions (MVD) concepts.

buildingSMART (2010) indicates that IDM aims to provide the integrated reference for process and the required data/information by identifying the discrete processes executed within building construction, the information required for their execution and the results of that activity. It specifies amongst other things the information created and consumed, and how the information should be supported by software solutions. A subset of the IFC Model Specification that is required for the information exchanges defined in one or more related IDMs is documented in an MVD (buildingSMART, 2012). This enables the implementation of those information exchanges in software products.

Design structure matrix / Dependency structure matrix (DSM)

Design Structure Matrix, also known as Dependency Structure Matrix (Gunawan, 2009; Zhao, Lv and You, 2008) serves as a tool to analyze sub-system relationships/interactions in complex engineering systems. There are four types of Design Structure Matrix: Product Architecture DSM, Organization Architecture DSM, Process-based Architecture DSM, and Multi-domain MDM, each serving to address respective problem type (Eppinger and Browning, 2012). DSM has been widely adopted in system engineering and engineering design because it effectively helps capture the essential exclusions or coupling relations of sub-system or components as well as activities and their dependencies. It can be used to represent architecture in terms of the relationships between its constituent components (Browning, 2001). The sequence of activities is vital in design projects, in particular interdisciplinary projects, and such sequencing is influenced by the information dependency among the activities (Maheswari and Varghese, 2005). Zhao, Lv and You (2008) applied DSM and Monte Carlo Simulation to predict change in construction project. Maheswari, Varghese and Sridharan (2006) identified DSM as a tool to represent information flows and dependency between activities, and applied it for activity sequencing in concurrent engineering projects. Numerical Design Structure Matrix (NDSM) derived from DSM could contain a multiple of attributes with more detailed information on the relationships between the different system elements (Gunawan, 2009).

RESEARCH METHODOLOGY

The study reported in this paper is part of a broader research of the Energy Efficient Buildings (EEB) Hub project (http://www.eebhub.org/). The materials utilized for the development of the information exchanges table originates as an extension from our current work on the Integrated Building Lifecycle Process Model (IBLPM), which is an extension of the earlier Integrated Building Lifecycle Process (IBLP) (Lee, 2012). This model was developed to visualize the building design process of various building elements (e.g., HVAC, lighting, structural, etc) based on OmniClass. The design process of each of the elements was modeled on a swimlane, which according to BMPN is a participant or actor who performs the process activities (OMG, 2011). We adapted this concept and specified the building elements in the swimlanes. The design process within the model is broken down into phases according to the Integrated Project Delivery (IPD) design phases (AIA, 2007). We used a range of methods to gather data including literature review, expert interviews, and validation by a workshop of a group of experts (Robson, 2011). We concentrated our efforts on studying and analyzing materials on construction design processes, where these were lacking or insufficient, construction industry experts and academics were approached as sources. We conducted a systematic review of the initially developed HVAC and structural process maps of the IBLPM which were adequately validated by the industry experts as the foundation in this work; to identify information items addressing 'information consumed' and 'information

produced' of each process activity. In some instances where we assumed an obvious information item/s was/were left out, we further reviewed literature and in some cases consulted experts to ascertain the necessity for its/their inclusion. The identified information was analyzed according to purpose, which helped in the definition of the categories and classes. The paper utilizes DSM concept as a major method of breaking down building elements information exchanges, and went on to focus on analyzing the swimlane-to-swimlane information exchanges. To do this, activity-based or process-based DSM used for modeling processes and activity networks and their information flow and other dependencies (Browning, 2001) was adopted.

INFORMATION EXCHANGES MAPPING

The information exchanges dependency table

The identification process of the information exchanges, from the activities developed in the process map, brought the formulation of an information exchanges dependency table. This table can be used to catalogue the information exchanges items based on several categorizations. The table's columns identify several properties of an information item. The structure of the information exchange table (IET), as we will refer to it for the rest of the paper, comprises the following components: *Swimlane; Phase; Activity; Information Consumed; Information Produced*. These are described as follows:

- Swimlane: This lists the individual building elements or systems (e.g., HVAC, Structural, etc).
- Phase: The different stages (i.e., conceptualization, criteria design, detailed design and implementation documents) of the design process are listed under this column.
- Activity: This column contains all the design steps or tasks that are executed during the process of designing a building element.
- Information Consumed: the various types of information required and used by the activities are identified and listed in this column.
- Information Produced: the various types of information generated as a result of executing a particular activity.

Development process

In order to develop the table, the initial step was examining each activity, and based on the authors' knowledge, information items were determined. In so doing, all the activities of each process were examined, without this, identification and categorization of the relevant information items would not have been possible. An important consideration was to be aware of the type of information required for and produced by each activity. For example, for the 'Conduct Airflow Analysis' activity in Figure 1, the following consumed information items were identified: 'energy and thermal zones model', 'required loads', 'initial HVAC system design' and 'HVAC performance parameters'. Additionally, the following information items were produced: 'HVAC system design' and 'HVAC parameters'. Based on this process of information identification, several observations were made. Each activity, in the process, can either consume and/or produce single or multiple information items as shown in Figure 1. Similarly, an information item can be consumed by multiple activities.

Categorization of information exchanges items

As mentioned earlier, the information exchanges items went through a process of categorization, sub-categorization, and classification. The information items were

categorized in the table's column based on their activity whether they were consumed or produced as the main category. After the identification of the information items was concluded, three sub-categories were developed to further distinguish them. These sub-categories differentiate the information items to be *product, performance or control* information. In addition, the information items are then classified into three classes based on their current status, namely: target*, predictive, and actual*. Table 1 tabulates the sub-categories and classes of the information exchanges along with their descriptions.

Table 1: Sub-categories and Classes of Information Exchanges Items

Sub-category	Description
Product	Product information exchange items areas related to the building's physical elements and properties, for example a wall height of 5ft.
Performance	Performance information exchange items pertain to information related to the building's performance (i.e., quality, cost, schedule, etc.). For example daylight factor of 2%.
Control	Control information exchange items are related to the information which influence the building design process. Control information can be either 'environmental' or 'process', for example, weather, cost, etc.
Classes	Description
Target (T)	Information that is used as a basis to evaluate and judge whether a design meets a desired outcome. That is the information which the project aims its efforts and eventual product to achieve.
Predictive (P)	Information that represents specific expected performance of a proposed designed. It specifies definitions and specifications of how the target information may be delivered during design.
Actual (A)	Information documented to measure the outcome of the proposed design.

The categorization was based on the purpose and use of an information item within the lifecycle phases. The sub-categories and classes can be embedded within the 'information consumed' and 'information produced' columns of the main information exchanges table (IET) (a sample demonstrated in Figure 1) when populating it with information exchanges items. The IET is populated based on the properties of each activity. The categories are color coded (*i.e., blue = 'product information' pink = 'performance information' and green = 'control information'*) in order to provide a visual assist to the user as shown in Figure 1. However, any other color could have been used for this purpose. Each of the classes has a related tag, (T) for target information class; (P) for predictive information class, and (A) for actual information class.

ANALYSIS AND DISCUSSION

The IET (Figure 1) demonstrates that not only product information but also several types are exchanged during design process. The information may come in different formats; from multiple sources and processed by different applications for energy simulation and analysis. Therefore, the table can serve as a database of information exchanges items, which can be a reference point for both analyst and software developers. It captures the information inputs of individual activities produced by other activities in the same or different swimlanes, and generates its own information. In this manner, if propagated across the lifecycle, the table will be able to comprehensively visualize the information flow among the activities. The table also shows that a comprehensive consideration is required to take into account the different types of information during energy simulation. From the general use concept of DSM, this consideration will contribute in effective coordination and integration efforts between

activities and simulation tools. Due to the significant number of iterative energy design simulation and analyses, the categorization will help in the compare and contrast of targeted and predictive information because the specific information would have been visualized accordingly. The table can support (automate) identification of relevant information for design change, product change and process change. Included in this is the identification of information constraints that affect planning, execution and coordination of the simulation and analysis activities, as well as other energy related activities. From technology and information exchange perspectives, the table can support tools to generate, track/trace, and consolidate relevant information that can be useful to other tools (where different tools are used to perform different activities). For example, some of the information items generated from an airflow analysis process using CONTAM can be transferred to EnergyPlus for use in energy analysis activity. The table can support identification and visualization of potentially useful simulation and analysis results of previous designs or even affected project phases that are relevant to a design, product or process changes.

HVAC	Criteria Design Phase	Select HVAC System Types	Architectural & Structural Model (P) HVAC Design Scope (T) HVAC System Requirements (T)	Initial HVAC System Design (T)
HVAC	Criteria Design Phase	Refine Spatial Requirements of the Systems	HVAC System Design (T) Architectural & Structural Models (P) HVAC Spatial Requirements (T)	HVAC Spatial Requirements (P)
HVAC	Criteria Design Phase	Zone Spaces into Systems	Architectural & Structural Model (P) HVAC Energy Parameters (T)	Energy Model (P)
HVAC	Criteria Design Phase	Outline HVAC Specs	HVAC System Design (T) HVAC System Requirements (T)	HVAC System Specs (T)
HVAC	Criteria Design Phase	Estimate Loads for Zones	HVAC Performance Parameters (T) Energy Model (P)	Required Loads (P)
HVAC	Criteria Design Phase	Conduct Airflow Analysis	Architectural & Structural Model (P) HVAC System Specs (T) Operations Schedule (P) Energy & Thermal Zones Model (P) Required Loads (P) Initial HVAC System Design (T) HVAC Performance Parameters (T)	HVAC System Design (P) HVAC Performance Parameters (P)

Figure 1: HVAC IET showing criteria design phase

Realizing that different users and activities produce different information and models with different software, and that information consumed by an activity is produced by other activities; there is great degree of change impact as a result of changes to an information item. Therefore, the table will help in understanding the effects of changing information items and sequence of activities across the design process, activities, phases and building elements, which is crucial for analyzing impact of the changes. This paper formulates a roadmap to utilize the IET to support development of IDM by specifying a link or interface between process maps and exchange requirements as shown on Figure 2. In this way, it is anticipated that the IET will enrich the identification of exchange requirements presenting in detail all the information exchange items during the integrated

design process. Identification of the information items holding the exchange requirements will become less laborious for domain experts.
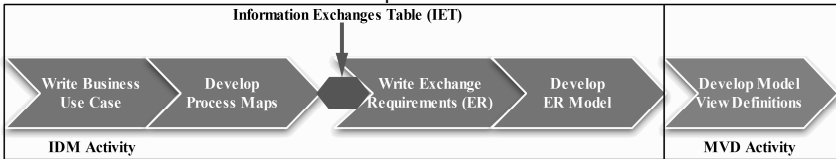
Figure 2: Context for utilization of the information exchanges table (IET) (Adapted from buildingSMART, 2012)

CONCLUSIONS

In this paper, we presented an Information Exchanges Table (IET) that can be used to document design information exchanges items, and HAVC IET to demonstrate its use. The table shows amongst other things information items which include both information consumed and produced, and the relating design activities of building elements. The table can be used as a resource reference point for project teams to understand activity-information dependencies during the design of building elements. This is beneficial in recognizing and considering how an information item is used by various activities, and/or how an activity can use multiple types of information. This creates information dependency, which is crucial to assess the impact of changing an information item in the design process. The paper also proposes an approach to utilize the table to support information exchanges initiatives. The limitations identified in this paper are that: (i) the information exchange table did not cover the entire lifecycle of the building but focused on only the design phase; (ii) definitions of the activities have not been included in the current version of the table. Our future work will include the development of information exchange tables for all other building elements/systems defined on the IBLPM, and to expand the table to include construction and operations phases. This will include the provision of descriptions or definitions of each of the activities. It is anticipated that this will provide better understanding of the meanings of the activities so that assumptions are not made about what they are. Furthermore, the table will be evaluated and validated with industry to determine its applicability in energy efficient retrofit projects, and general construction.

ACKNOWLEDGEMENT

REFERENCES

AIA. (2007). "Integrated Project Delivery: A Guide".
  <http://www.aia.org/groups/aia/documents/pdf/aiab083423.pdf > (March 2012).
Brewer, G., and Gajendran, T. (2009). "Emerging ICT trends in construction project teams: a Delphi survey." *Journal of Information Technology in Construction,* 14, pp. 81 – 97.
Browning, T.R. (2001). "Applying the Design Structure Matrix to System Decomposition and Integration Problems: A Review and New Directions." *IEEE Transactions on Engineering Management*, 48(3), pp.292–306.

buildingSMART. (2010). "Information Delivery Manual guide to Components and Development Methods." <http://iug.buildingsmart.org/idms/development/IDMC_004_1_2.pdf/view> (December 2012).

buildingSMART. (2012). "An Integrated Process for Delivering IFC Based Data Exchange." <http://iug.buildingsmart.org/idms/methods-and-guides/Integrated_IDM-MVD_ProcessFormats_14.pdf/view> (Accessed November 2012).

Eppinger S. D., Browning T.R. (2012). *Design Structure Matrix Methods and Applications*. Boston: MIT Press.

Geryville, H. M., Ouzrout, Y., Bouras, A. and Sapidis, N. S.(2005). "A collaborative framework to exchange and share product information within a supply chain context". In *Proceeding of IEEE, International Conference on Machine Intelligence*, November 4-7, Tozeur, Tunisia, pp. 195-202.

Gunawan, I. (2009). "Application of Numerical Design Structure Matrix Method in Engineering Projects Management." *Operations and Supply Chain Management*, 2(1), pp.1–10.

Han, Z. and Froese, T. (2009). "Project Information Management in Mega Oil Sands Projects." Building a Sustainable Future: Proceedings of the 2009 Construction Research Congress, April 5-7, Seattle, Washington, ASCE, pp. 71-80.

Hjelseth, E. (2010). "Exchange of Relevant Information in BIM Objects Defined by the Role and Life-Cycle Information Model." *Architectural Engineering and Design Management*, 6, pp. 279-287.

Kriphal, M. and Grilo, A. (2012). "Compatibility between design and construction building information models." In *eWork and eBusiness in Architecture, Engineering and Construction*. CRC Press, pp. 447–452.

Lee, S., Liu, Y., Chunduri, S., Solnosky, R., Messner, J., Leicht, R. and Anumba, C. J. (2012). "Development of a Process Model to Support Integrated Design for Energy Efficient Buildings." In: Issa, R. and Flood, I. (eds) The Proceedings of the ASCE International Conference on Computing in Civil Engineering, June 17-20, Clearwater beach, florida, USA, pp. 261-268.

Maheswari, J.M. and Varghese, K. (2005). "A Structured Approach to Form Dependency Structure Matrix for Construction Projects." In *Proceeding of the 22nd International Symposium on Automation and Robotics in Construction*. September 11-14, Ferrara, Italy, pp. 1–6.

Maheswari, J.M., Varghese, K. and Sridharan, T. (2006). "Application of Dependency Structure Matrix for Activity Sequencing in Concurrent Engineering projects." *Journal of Construction Engineering and Management*, 132(5), pp.482–490.

OMG. (2011). "Business Process Model and Notation (BPMN) Version 2.0." < http://www.omg.org/spec/BPMN/2.0> (Accessed December 2012).

Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*. 3rd (ed). Chichester: Wiley.

Sanvido, V. E., Anzola, G., Bennett, S., Cummings, D., Hanlon, E., Kuntz, K., Lynch, T., Messner, J., O'Connor, E., Potter, K., Riley, D. and Yoshigi, T. (1995). "A Process Based Information Architecture." Technical Report No. 36, Computer Integrated Construction Research Program, The Pennsylvania State University.

Zhao, Z., Lv, Q. and You, W. (2008). "Applying Dependency Structure Matrix and Monte Carlo Simulation to Predict Change in Construction Project." In *Seventh International Conference on Machine learning and Cybernetics*. Kunming, China: IEEE, pp. 670–675.

Calculating the Cost of Heating and Cooling Loss for Building Diagnostics using EPAR - Energy Performance Augmented Reality- Models

Youngjib Ham[1] and Mani Golparvar-Fard[2]

[1] Ph.D. Student, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801; PH (217) 259-5498; FAX (217) 265-8039; email: yham4@illinois.edu

[2] Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801; PH (217) 417-9552; FAX (217) 265-8039; email: mgolpar@illinois.edu

ABSTRACT

Successful building diagnostics requires quick and reliable identification of energy performance problems, and an accurate assessment and cost analysis of the associated energy loss. Today, most auditors rely on thermal imagery for analyzing energy performance problems. Nonetheless, current thermographic inspections are qualitative and thus mainly focus on the visual detection of thermal irregularities. As a result, the quality of the retrofit decision-makings is based on how these images are interpreted by the auditors. To overcome these challenges, this paper presents a new method for calculating the cost associated with potential building energy performance problems. First, based on the Energy Performance Augmented Reality (EPAR) modeling method, actual and expected 3D spatio-thermal models of the buildings under inspection are generated and integrated in a 3D environment. This method leverages thermal and digital images captured by a thermal camera for actual energy performance modeling, in addition to computational fluid dynamics (CFD) analysis for expected energy performance simulation. Based on thermal 3D mesh modeling and threshold-based assessments of temperature deviations, building areas with potential performance problems are detected within the EPAR models. By locating a small crumbled aluminum foil on an inspection surface, the reflected apparent temperature is measured. For areas with potential performance problems within EPAR models, this measurement is used to calculate the actual R-value at the level of 3D points. This information is finally used to calculate both heat loss and gain and also estimate the associated energy costs. The presented method is tested on several interior locations of an existing residential building. Experimental results and the benefits of converting temperature data sensed from building surfaces into actual R-values and the cost of heating and cooling loss are discussed.

INTRODUCTION

With the emerging energy crisis and recent legislative measures on energy conservation, the building industry is paying more attention to decreasing the level of energy demand in new and existing buildings. For example, practitioners are more often using building materials with high thermal resistance for insulation to help

homeowners lower the cost for heating and cooling. Nonetheless, according to a recent report by the U.S. Department of Energy (U.S. DOE 2010), around 35% of input energy is still wasted by the building sector. Among them, building performance problems account for over $80 billion of wasted value per every year. Despite its significance, identifying building performance problems and improving energy efficiency is not a trivial task. Currently, building energy performance simulation tools such as EnergyPlus and Ecotect are widely used for the purpose of building retrofit decision-making. However, these simulation tools have limitations on finding building areas with performance problems. For example, they mainly focus on analyzing building energy loads such as annual energy consumption for comparing different retrofit alternatives and varying patterns of building occupancy.

In the meantime, infrared thermography – which detects and measures heat variations from building surfaces – is considered as an emerging non-invasive inspection tool for the diagnostics of poor building conditions (Cerdeira et al. 2011). Analyzing thermal performance data collected from thermal cameras helps in locating energy inefficiencies such as poor insulation and air leakages. Ultimately, it can help save energy needed for heating and cooling by improving building energy efficiency. Nonetheless, the current practice for identifying building inefficiency and saving the associated energy cost using thermography has the following challenges: 1) since a large number of unordered thermal images needs to be examined manually, their direct application for the entirety of building diagnostics is time-consuming and labor-intensive; 2) the auditors' interpretations of thermal images are also subjective and often inconsistent. Considering the limited number of trained professional auditors, their decisions may adversely affect the quality of building diagnostics. Without reliable benchmarks for energy performance, the analysis and interpretation of thermal images –even for trained professional auditors– could be error-prone. Finally, 3) current practices for thermographic inspection are mainly performed for qualitative analysis; i.e., they mainly focus on the visual detection of abnormal thermal regions such as hot and cold spots. For building retrofit, most homeowners are typically motivated by how much money can be saved if the observed problems are remedied. Without such quantitative interpretations of building thermal images, homeowners may be reluctant to invest their money for retrofitting their buildings due to the high degree of uncertainty in the value of their investment.

Recently, we presented Energy Performance Augmented Reality (EPAR) modeling method to address the challenges associated with subjective and manual interpretations of large numbers of unordered thermal imagery (Ham and Golparvar-Fard 2013). Building on prior research on EPAR models, this paper presents a new method for calculating the energy cost associated with heating and cooling loss by measuring the actual R-value of building materials at the level of 3D points. This method can quantify the energy and monetary loss associated with potential performance problems and in turn motivate homeowners to retrofit their buildings by systematically analyzing the Return on Investment (ROI) for various retrofit alternatives. In the following sections, we first review the related works in measuring R-value and energy loss. Next, the underlying assumptions and algorithms for calculating heat loss/gain and the associated costs are presented in detail. Finally, the experimental results, perceived benefits, and challenges are discussed.

BACKGROUND AND RELATED WORKS

R-value of building environments. *R-value* is the measure of thermal resistance of a building material or element to the heat flow (Desogus et al. 2011). Higher R-value in building materials indicates better insulations and a greater potential for building energy savings. During the operational phase of a building due to building structural problems such as degradation of old insulations, the ability of building materials to resist heat flow typically decreases. As a result, the actual R-values for building elements will be lower than the notional values declared by their manufactures or those estimated during the design phase (Desogus et al. 2011). For energy efficient buildings, there is a need for 1) using building materials with high R-value; and 2) monitoring the actual R-value of building elements during the operational phase.

Measuring actual R-values of building elements using thermography. Currently, there are two common practices for in-situ measurement of R-value for building environments: 1) *destructive method*: analyzing building component layers through direct measurement; and 2) *non-destructive method*: measuring the heat-flux between inside and outside building such as (Cabeza et al. 2010). For non-destructive measurements, a few researchers have focused on application of thermography. These methods measure the actual R-value of building envelops to verify building insulation. Madding (Madding 2008) developed an algorithm for measuring the R-value of drywall sections within stud frames using thermal images. This work estimated energy savings based on various insulation retrofit alternatives. Similarly, Fokaides and Kalogirou (Fokaides and Kalogirou 2011) measured the overall heat transfer coefficient (U-value which is the reciprocal of R-value) for building envelops using thermal images. This work focused on in-situ measurement of U-values and compared them with their notional values. It also validated thermography for evaluating performance of building elements with respect to their resistance to heat transfer. These methods assume that each inspected area has a constant R-value. Thus, they cannot accurately represent the actual R-value variations which are caused by non-uniform deteriorations in building elements. Furthermore, these methods require a large number of thermal images to be manually explored to measure the actual R-value for each building element. As a result, their application for the entirety of a building can be time-consuming and labor-intensive. There is a need for more robust methods using thermal cameras that can rapidly and reliably explore the actual heat transfer conditions associated with potential performance problems.

OVERVIEW OF THE PROPOSED METHOD

High thermal resistance reduces the total energy consumption required for heating and cooling and enables building occupants to have an optimal thermal comfort at a lower energy cost. In steady-state conditions, the overall heat transfer (Q) rate through a wall surface (A) with a temperature difference between inside and outside (ΔT) can be described using the following equation:

$$\frac{dQ}{dt} = \frac{1}{R} \times A \times \Delta T \tag{1}$$

In the proposed method as shown in Figure 1, the EPAR modeling method is used to automatically identifying the building areas (*A*) that have potential performance problems. Next, the actual R-values of all inspected building surfaces are measured at the level of 3D points. Finally, using Eq. 1, the heat loss/gain and the associated energy costs are calculated. The following sections describe each step in detail:
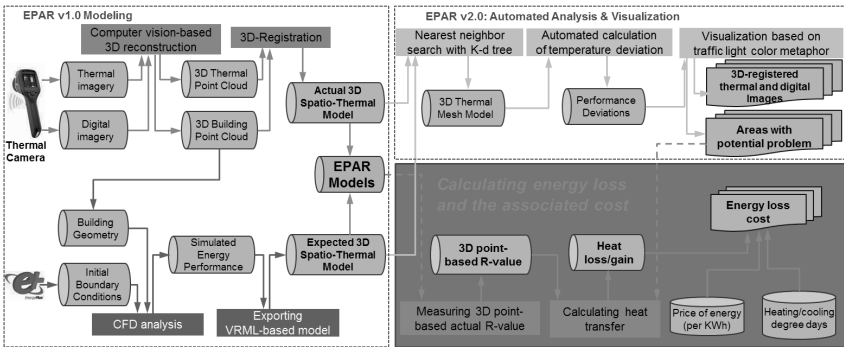
Figure 1. Overview of data and process in the proposed method

Identifying building areas with potential performance problems for retrofit. Our method for calculating the cost associated with heating and cooling loss for retrofit analysis builds upon the recently prototyped EPAR models (Ham and Golparvar-Fard 2013). For EPAR modeling, we leveraged the following two emerging sources of information: 1) unordered collection of digital and thermal images: these images are collected using a single hand-held thermal camera with a built-in digital camera and enable the actual energy performance of the building under inspection to be modeled in form of 3D spatio-thermal point cloud models (Ham and Golparvar-Fard 2012); 2) VRML based-computational fluid dynamics (CFD) models: CFD simulation provides the detailed spatial distributions of building thermal performances and therefore enables the expected energy performance to be modeled in 3D. Recently, using *k-d* tree structure and nearest neighbor searching, we generated thermal 3D mesh models for the best representation of the actual thermal performance. Through a systematic comparison of the measured and simulated building energy performances in 3D mesh-based EPAR models, performance deviations are calculated and analyzed based on the thresholds for identifying potential performance problems. Finally, the detected potential performance problems are intuitively visualized in the EPAR models using a metaphor based on traffic light colors. The readers are encouraged to look into (Ham and Golparvar-Fard 2013) for more details on the process for creating EPAR models and identifying potential performance problems in the EPAR models. Next, for calculating building areas with potential performance problems, we compute the surface area of the generated triangle thermal mesh in the EPAR models. We first find the faces from the meshes which are associated with potential performance problems. By using the three vertices which form each face, we then calculate the areas using the cross product of the two corresponding vectors, and aggregate them. Figure 2

summarizes the algorithm for calculating the surface area of meshes with potential performance problem (A_P).

Input: Three vertices in thermal triangle mesh

$$Q_1 = \{P_1^i | \forall i \in (1,2,\dots,l), P_1^i =< X_1^i, RGB_1^i >\}$$
$$Q_2 = \{P_2^i | \forall i \in (1,2,\dots,l), P_2^i =< X_2^i, RGB_2^i >\}$$
$$Q_3 = \{P_3^i | \forall i \in (1,2,\dots,l), P_3^i =< X_3^i, RGB_3^i >\}$$

Output: Area with potential performance problem (A_P)

1　**for** i=1:l
2　　**if** RGB_1^i & RGB_2^i & $RGB_3^i = (255, 0, 0)$
3　　　return X_1^i, X_2^i, X_3^i
4　　**end if**
5　　$A_P^i = \dfrac{\left|X_2^i X_3^i - X_3^i X_2^i, X_3^i X_1^i - X_1^i X_3^i, X_1^i X_2^i - X_2^i X_1^i\right|}{2}$
6　**end for**
7　return $\sum A_P^i$

Figure 2. Calculating building areas with potential performance problems in the EPAR models.

Measuring the actual R-values of building surfaces at the level of 3D points. Our work for estimating the R-value is based on the assumption that the heat transfer between building materials and the thermal camera lens is attributed to the following thermal convection and radiation equations:

$$Q_{Con} = \alpha_{con} \times Area \times \left(T_{inside,air} - T_{inside,wall}\right) \tag{2}$$

$$Q_{Rad} = \varepsilon \times \sigma \times Area \times \left(T_{inside,wall}^4 - T_{inside,refleted}^4\right) \tag{3}$$

Where α_{con} is the convective heat transfer coefficient, ε is thermal emissivity, σ is Stefan-Boltzmann constant (5.67×10^{-8} W($m^2 K^4$)$^{-1}$). By combining Eqs. 1, 2, and 3, the total thermal resistance can be calculated using the following equation:

$$R = \frac{T_{inside,air} - T_{outside,air}}{\alpha_{con} \times \left(T_{inside,air} - T_{inside,wall}\right) + \varepsilon \times \sigma \times \left(T_{inside,wall}^4 - T_{inside,refleted}^4\right)} \tag{4}$$

Within the EPAR models, $T_{inside,wall}$ can be queried at the level of 3D point. $T_{inside,reflected}$ is the inner surface temperature of building environments, which is assumed to act as a large thermal reservoir without temperature change. To measure this value, we used a small crumbled aluminum foil located on the inspection areas (Figure 3) (FLIR system 2010). Since the foil has low emissivity and high reflectivity, we can robustly measure the reflected apparent temperature from the
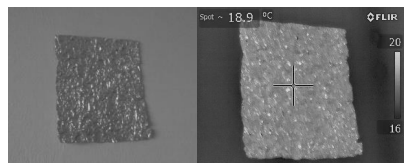
Figure 3. Measuring the reflected apparent temperature

crumbled foil. α_{con} is subject to airflow types such as laminar or turbulent and temperature deviations between building surfaces and the air. In this paper, we adopted the convective heat transfer coefficient from (ISO 6946:2007). $T_{inside,air}$ and $T_{outside,air}$ are measured using a thermometer. Ultimately, by using Eq. 4 based on the EPAR model, actual R-value of building environment can be calculated at the level of 3D point.

Calculating heat gain/loss and the associated costs. The heat loss/gain associated with the potential performance problems (Q) are calculated using Eq. 1 based on the following three values: 1) detected building areas with potential performance problems in the EPAR models (A_P); 2) measured actual R-values for 3D points; 3) '*degree days*' which is the combination of the time (t) and temperature difference between the average outdoor and a predefined baseline (ΔT). (The National Oceanic and Atmospheric Administration (NOAA) 2013). The final step of our work is to estimate the energy cost associated with the calculated heat loss/gain by combining with the current retail price of energy such as electricity or gas.

EXPERIMENTAL RESULTS

Data collection and experimental setup. To validate the proposed method, several case studies were conducted on the interior of an existing residential building in the Commonwealth of Virginia. For the EPAR modeling, digital (2048×1536 pixels) and thermal (320×240 pixels) images were captured by using an FLIR E60 thermal camera which has a built-in digital camera. Gambit 2.2.30 and Fluent 6.2.16 are used for simulation model-setup and CFD analysis respectively. The detailed experimental setups for CFD analysis are summarized in Table 1.

Table 1. Initial boundary conditions for CFD simulation

Contents	Value
Viscous Model	RNG k-ε model
Number of vertices in mesh	2,830,660
Fluid (air) motion type	Stationary
Temperature of Inlet Air	19°C (Cooling)
Velocity of Inlet Air	0.6 m/s
Wall motion type	Stationary
Wall roughness height	0
Wall roughness constant	0.5
Wall shear condition	No slip
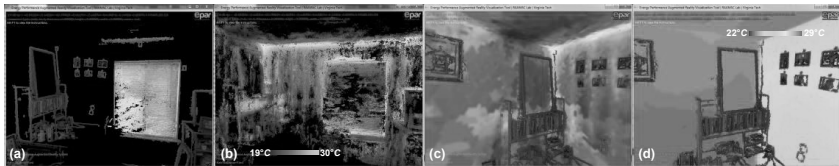
Figure 4. (a): 3D building geometrical point cloud, (b): 3D thermal point cloud, (c): 3D thermal mesh model, (d): VRML-based CFD model.

Results. Figure 4(a) and (b) show 3D building geometrical and thermal point cloud models respectively. Figure 4(c) and (d) present the 3D thermal mesh model and the

VRML-based CFD model. As observed, even for the areas that are sparsely reconstructed in point cloud model (Figure 4b), the 3D thermal mesh model (Figure 4c) provides detailed information on the actual thermal performance. Figure 5(a) and (b) show examples of the building areas with the detected potential performance problems in the EPAR models. The areas highlighted with red are originated from the deviations between the actual measurement from thermal images (Figure 4c) and the simulation results from CFD analysis (Figure 4d) in form of thermal 3D mesh models. As a proof of concept, considering the measurement accuracy of the FLIR E60 camera and typical accuracy of the CFD simulations for modeling indoor environments, the deviations above the 2˚C threshold are detected in this paper. Thermal deviations that are above the predefined threshold provide the auditors with feedbacks on those areas that require additional detailed performance analysis. Here, the detected areas show the typical degradation of thermal resistance (5a): between a side wall and a ceiling around the HVAC system and (5b): between a side wall and a floor adjacent to building exterior. Figure 5(c) highlights an observed problem using an automatically 3D-registered thermal image. Through visual inspection of the 3D-registered thermal image in the EPAR models, the detected potential performance problems can be qualitatively validated. Figure 5(d) illustrates a 3D-registered digital image which visually represents the building semantics of the same areas captured in the thermal image shown in 5(c). The figure is best seen in color.
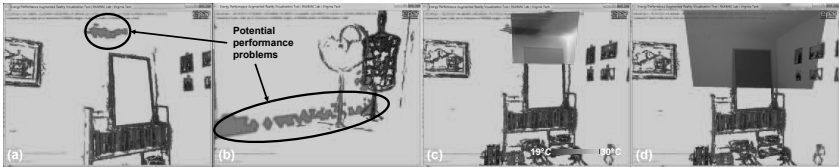
Figure 5. Potential performance problems along with a 3D building geometrical point cloud and automatically 3D-registered images in EPAR V2.0 models.

Table 2 reports the calculated building areas with potential performance problems, the heat loss/gain, and the associated energy cost in our experiment. As a proof of concept in this paper, for calculating the cost of energy loss associated with annual cooling loss (i.e., heat gain) during the summer, we used 'cooling degree days' data of Virginia. As a result, we calculated the cost of the cooling loss for this room to be $10.06 per year. Ultimately, converting the temperature data sensed from building surfaces into the cost of heating and cooling loss can help building owners understand current building energy efficiency and analyze the cost/benefit ratio for various retrofit investments.

Table 2. Calculating heat loss/gain and associated energy cost

Areas with potential problems (m^2)	Heat loss/gain (J/S)	Price of electricity[+] (cents/kWh)	Degree days[++]	Energy loss cost ($)
0.94	2.82	11.22	1326	10.06

[+] U.S. Energy Information Administration (2012)
[++] The National Oceanic and Atmospheric Administration (2013)

CONCLUSIONS AND FUTURE WORKS

Current thermographic inspections mainly focus on qualitative interpretation of a large number of unordered 2D thermal images and locating thermal anomalies based on an auditor's knowledge. To reliably and rapidly identify potential performance problems from the thermal images and quantify their energy loss and monetary impact, we present a new method for calculating the heating loss/gain and the associated energy cost by measuring the actual R-value of building environments within EPAR models at the level of 3D points. The quick and reliable assessment of the cost benefit ratio for retrofit investments in our proposed method can motivate homeowners to consider retrofitting their buildings. Our future works involve implementing extensive experiments, creating an algorithm for application of different emissivity of diverse building materials, and elimination of false positives from the detected performance problems. These areas are currently being explored as part of our ongoing research.

REFERENCES

Cabeza, L. F., Castell, A., Medrano, M., Martorell, I., Pérez, G., and Fernández, I. (2010). "Experimental study on the performance of insulation materials in Mediterranean construction." *Energy and Buildings*, 42(5), 630-636.

Cerdeira, F., Vázquez, M. E., Collazo, J., and Granada, E. (2011). "Applicability of infrared thermography to the study of the behaviour of stone panels as building envelopes." *Energy and Buildings*, 43(8), 1845-1851.

Desogus, G., Mura, S., and Ricciu, R. (2011). "Comparing different approaches to in situ measurement of building components thermal resistance." *Energy and Buildings*, 43(10), 2613-2620.

FLIR system (2010). "Thermographhic measurement techniques." *FLIR Reporter Professional, Professional Edition*, 133-137.

Fokaides, P. A., and Kalogirou, S. A. (2011). "Application of infrared thermography for the determination of the overall heat transfer coefficient (U-Value) in building envelopes." *Applied Energy*, 88(12), 4358-4365.

Ham, Y., and Golparvar-Fard, M. "Rapid 3D Energy Performance Modeling of Existing Buildings using Thermal and Digital Imagery." *Proc., 2012 Construction Research Congress*, 991-1000.

Ham, Y., and Golparvar-Fard, M. (2013). "EPAR: Energy Performance Augmented Reality Models for Identification of Building Energy Performance Deviations between Actual Measurements and Simulation Results." *Energy and Buildings*.

ISO 6946:2007 "Building components and buildin elements-Thermal resistance and thermal transmittance-Calculation method."

Madding, R. "Finding R-Values of Stud Frame Constructed Houses with IR Thermography." *Proc., InfraMation 2008*, 261-277.

The National Oceanic and Atmospheric Administration (NOAA) (2013). "Heating & Cooling Degree Day Data." <http://www.ncdc.noaa.gov/oa/documentlibrary/hcs/hcs.html>.

U.S. Department of Energy (2010). "2010 U.S. DOE buildings energy databook."

A Cross-Case Analysis of Decision Making Environment for Deep Retrofit Projects

Pelin Gultekin[1], Chimay J. Anumba[2] and Robert M. Leicht[3]

[1]Graduate Research Assistant, Department of Architectural Engineering, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802.
e-mail: pug118@psu.edu
[2]Professor, Department of Architectural Engineering, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802.
[3]Assistant Professor, Department of Architectural Engineering, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802.

ABSTRACT

Retrofitting existing buildings has the greatest opportunity to reduce the energy consumption in the building market. Even though the opportunity lies in making these design decisions collaboratively, it is difficult to come up with "best project fit" with current techniques by considering all disciplines' priorities. Especially, deep retrofits that provide higher energy targets require integrated design approaches that include the perspectives and priorities of the various disciplines through collaborative decision making within the limits available budget. The aim of this project is to define the critical decisions - when, why, and how these decisions are made in contrast to other design options - by the decision makers in the team. Two deep retrofit case studies will be presented and will deliver integrated process maps and comparative analysis using different delivery methods. The maps and decision criteria are being used in the development of a Decision Support System (DSS) aimed at providing guidance to retrofit project teams by increasing the transparency of the decision making process and highlighting the energy efficiency implications of design alternatives.

INTRODUCTION

Decision making in a collaborative environment is always challenging, especially when different alternatives and attributes are involved. The process involves designers and decision makers that have several interests in developing their own ideas that satisfy their own objectives. To solve this issue, research has been developed in collaborative design decision making to eliminate the uncertainty involved in the process. The main goal of this work is to design a methodology involving the creation and selection of design objectives to assess the accuracy in the selection process of the best alternative in a collaborative environment involving a limited number of decision makers.

Delivery processes, organizational team structures, and acquisition strategies for team members have an elevated importance on high performance green projects due to the increased number of user groups, design consultants and specialty contractors that must interact during the design and construction. Practitioners and academics have long been aware that the greater project integration in the design and construction process leads to better project outcomes in terms of cost, schedule and quality. More recent research has shown that these critical concepts that are decided at conceptualization phase are also correlated with high performance green project goals (Gultekin et al. 2012). Collaboration at earlier phases is becoming more important than ever to achieve a high performance design and meet the design targets after construction.

Retrofit projects specifically require more integration due to need for resolution of the inherent characteristics of the existing building and designing best fit project for these specific characteristics. The performance goal is always shaped in the limits of the existing conditions. Hence, designing the best fit project is more important than designing with the highest energy goals. The unique inherent characteristics add more constraints to be considered than a traditional green building project. For an integrated design approach, the key participants should become involved early and spend more effort on the design than in the conventional approach. Design charrettes are an extremely valuable sub-phases to integrate the team for subsequent phases.

This study will use process mapping methodology to capture the retrofit design process. The first vital step to understand the process is choosing a methodology for modeling which will be able to provide a foundation for clearly communicating the process, steps, information flow, and decision making for the project. The chosen methodology will be used to delineate the building EIC (Connelley School) and EEB Hub (Building 661) processes and provide clear documentation to allow the transferable techniques from these projects to be used to deliver future retrofit projects.

RELATED WORK

There are many ongoing efforts to understand and improve the delivery processes of buildings. As building performance improves, the design and construction processes become more inclusive and complex in terms of the involvement of various disciplines and discussion of multi-dimensional issues early on the project. Process modeling has been developed as a consequence of need for

understanding the articulated steps specifically, and analyzing the relationship of sequenced steps.

1. Decision Making Environment of Deep Retrofits

The review of related work is limited, since the retrofitting experiences in the level of high performance energy efficient deliveries are currently evolving in regards to new technologies and processes followed. The Rocky Mountain Institute delivered a comprehensive study on differentiating activities required for delivering deep retrofit projects. This study is utilized to establish the boundaries of design process understanding regarding design activity sequences. Incorporating energy efficiency related decisions into design at earlier phases is crucial. Hence, we converted the recommendations that are stated by RMI into a "to be" process map. "Early Design" phase is the earliest phase of a proposed design process (Figure 1). It has unique steps as setting measurable goal for energy reduction performance, gathering data from previous utility bills, audit results, etc., and calibrated energy model with inherent building conditions to develop a baseline prior to start design (RMI, 2012).
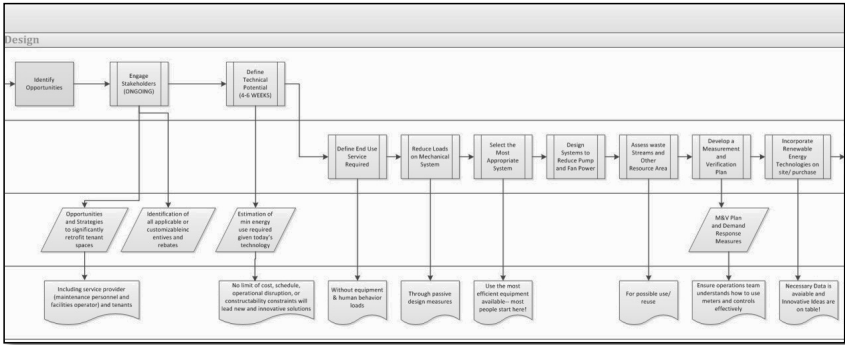
Figure 1: Early design phase recommendations adapted from RMI, 2012

2. Case Studies

This study includes the results of two case studies, Connelly School Retrofit Project and Navy Yard Building 661 Retrofit Project in Pennsylvania. Both projects are superior examples regarding the collaborative delivery process and energy efficient technologies used. The decision making is transparent due to the aim of delivering showcase projects in terms of energy efficient technologies and techniques used as well as the process followed.

2.1. Connelly School Retrofit Project, Pittsburgh, PA

Connelly School retrofit project (300,000sqf) is a green technology demonstration showcase building under the ownership of Pittsburgh Green Innovators Inc. This deep retrofit project achieves almost 49% reduction in energy consumption of the building through the integrated mechanical, electrical and plumbing (MEP) system design decisions. Decision making environment was collaborative due to the design-build delivery type selection.

The building typology is in two forms and functions; saw-tooth roofed section serves as a light industrial/manufacturing function and the tower portion serves for

certain tenants related to green technology suppliers with training purposes. These two sections will be equipped with different technologies in order to observe the performance levels for learning and demonstration purposes. The building systems is designed to demonstrate demand side energy management strategies.

2.2. Building 661 Retrofit Project in Navy Yard, Philadelphia, PA

The Building 661 project part of is newly named Energy Efficient Buildings (EEB) Hub; a 36,648sqf research laboratory and education facility. It is established in Philadelphia by the U.S. It is an existing two story building constructed in 1942 as a recreational building with three principal volumes: a two story 'headhouse' that housed locker rooms, restrooms, offices, and a bowling alley and two 'highbay' spaces; one of which housed a swimming pool and the other a basketball court. The Department of Energy (DOE) has a unique dual mission of improving energy efficiency in buildings and promoting regional economic growth and job creation from our headquarters in Philadelphia's Navy Yard (2012). Even though the delivery type is defined as Design-Bid- Build, the project team is working on an integrated design process non-contractually.

The building will be divided into three distinct zones where system cost, operation, and performance can be both viewed by the public and compared within the building. The building area systems will operate independently of each other for cooling, but will share heating elements connected to a hot water boiler system. Area 1 will be served by a chiller for cooling, with dedicated outdoor forced air/dew point control system with chilled beam terminal cooling and radiator hot water heating. Area 2 will be packaged direct expansion (DX) rooftop units with displacement ventilation for the large meeting rooms and variable air volume (VAV) to the common areas for cooling. Area 2 heating will be through radiator hot water heating or hot water coils in the VAV boxes. Area 3 will be ductless heat pumps for heating and cooling connected to a single outdoor unit, with hot water radiators as back-up heat on the coldest days. Area 3 will utilize natural ventilation to meet the ventilation.

3. Units of Analysis

The first step to evaluate a performance or process is relying on developing reliable measures that is evaluated via criteria. The developed criteria for evaluating the captured processes after a clearly defined problem statement set the evaluation basis in order to select the optimum decision compared to other design decision options (Dan 2004). Hence, in this study we are tracking the critical design decisions criteria via process models firstly for setting the basis of design options.
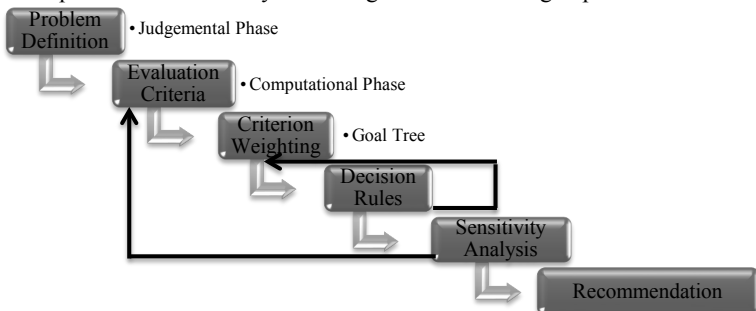
Figure 2: Decision Making Process (Adapted from Dan 2004)

Even though we are also tracking thermal, indoor air quality measures, we will limit the analysis in the scope of energy efficiency and cost measures regarding the performance outcome. In the light of the review, we summed up the units of analysis for energy efficiency as below:

- LEED Scorecard
- Normalized Annual Energy Consumption and Energy Use for heating in kWh/m2 (Rey, 2004; Zhu, 2006)
- Annual Electricity Use in kWh/m2 (Rey, 2004)
- Energy and Time consumption Index (ETI) (Chen et al, 2006)
- Energy Saving by Retrofitting expressed by (1- Energy/ Energy Baseline)% (Gholap and Khan, 2007)

For cost evaluation units of analysis:

- Total Cost of Ownership (Gartner, 1987)
- Direct Cost and Initial Investment Costs (Rosenfeld and Shohet, 1999)
- Annual Ongoing Maintenance Charges (Rosenfeld and Shohet, 1999; Rey, 2004)
- Annual Ongoing Charges (Rey, 2004)
- Net Present Value (NPV) of the Energy Investment (Martinaitis et al, 2004)
- Internal Rate of Return (IRR) of the Energy Investment (Martinaitis et al, 2004)

Another analysis criterion is building process services and management strategies developed by the project team. These measures are used to highlight the project values and goals in order to achieve targeted performance. Parallel to literature review of criteria used in energy efficient building industry, we run semi-structured interviews, personal meetings, group meetings, and follow-up meetings questions/documents using open-ended questionnaires are conducted to the project participants who joined design decision making process. What we are asking for is what the critical decisions are, how-why these decisions are taken in contract to other options at what time, and by whom (Gultekin et al. 2012).

4. Process Modeling

The building processes are tracked and documented with the intent to explain, allow, and reflect the collaborative environment of design and construction for deep retrofits. The "as-is" process model map and accompanying description will both improve the understanding of the process and the team's integration for the performance target of 50% improvement in energy efficiency. This descriptive model will also provide information on the interactions, information flow, and decision making within the project team; resulting in a better understanding of the process for deep renovations building projects. Several modeling techniques have been developed to reflect dynamic interaction. Quality of the models differs based on who and how it will be used. Also, discussions are ongoing to evaluate mainly usefulness and correctness of models in depicting processes (Hommes 2004). More than these two general aspects, a detailed study evaluated business process models with eleven criteria (Figure 3).
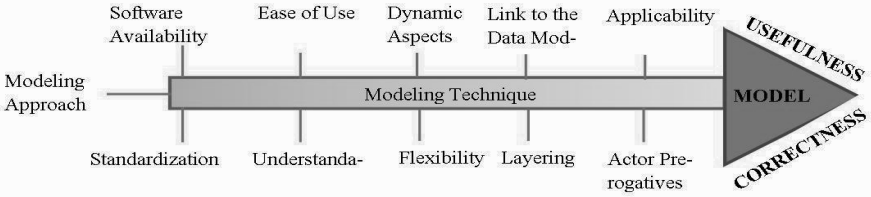
Figure 3: Evaluation Indicators for Business Process Modeling

Selection of Business Process Modeling Notation

Even though there are several business modeling techniques, it is critical to select the business model based on its usefulness for the specific service of interest and its correctness in reflecting the real life application of the model. Besides that, there are common benefits which can express and support the future view of business services. Some of them can be listed as below (C.J. Anumba 1998): Clearly specify people's role and tasks in the organization; provide accurate requirements for a subsequent information system supporting the business; model level benchmarking-i.e. trying new/ different business concepts at the model level and study the results.

Comparison of selected business modeling techniques was done based on the previously explained criteria. Characteristics of IDEF, DFD, RAD, and UML have been adopted from a comparative research (C.J. Anumba 1998). BPMN, a recent technique, has been reviewed and adopted to the same framework, shown in Table 1. Briefly, Business Process Modeling Notation (BPMN) (Martin Owen 2011) is used to model the integrated process for interdisciplinary teams to comprehend, communicate, and contribute in its explicit detailed structure.

Table 1: Comparison of Business Modeling Techniques (C.J. Anumba 1998)

Criteria	IDEF0	DFD	RAD				UML	BPMN
Modeling Approach	Static Activities	Data Flow Diagram	Emphasis on Roles: Sequence of actions and interactions				Object-oriented(OO)	Flowchart technique: Process-centric
Software Availability	Yes	Yes	Yes	Yes			Yes	
Standardization				Yes			Yes	
Ease of Use	Yes	Yes	Yes	Yes/ No			Yes	
Understandability	Yes	Yes	Fair	Yes/No			Yes	
Dynamic Aspects	No	No	Yes	Yes			Yes	
Flexibility	Fair	Fair	Fair	Yes			Yes	
Link to the Data Model	Yes	Yes	No	Yes			Yes	
Layering	Yes	Yes	No	No			Yes	

Applicability	Functional Modeling	Data Flows	Software Process Modeling	OO Analysis and Programming	**Business and Software Process Modeling**
Actor Prerogatives	Limited (Mechanism)	No	Yes	Yes	**Yes**

CONCLUSION

Two projects, EEB and EIC are superior and comprehensive examples in green building industry. Not only building performance levels but also process improvements are upgrading the practices regarding efficiency. In this study, we focused on building systems, and building process services and management decisions criteria followed through the data collection process. Performance indicators for systems and process are listed as below:

Table 2: Comparison of the criteria used in two projects

Comparison Criteria	EEB	EIC
Energy Efficiency		
LEED Scorecard	Gold	Platinum
Energy Star Rating	82	N/A
Energy Saving by Retrofitting	44%	49%
Annual Energy Consumption*	35%	45%
Embodied Energy	Informally considered	Not Considered
Energy Use Intensity (EUI)	63kBtu/sf-year	N/A
Cost Efficiency		
Annual Energy Cost*	24%	N/A
First cost of technologies	✓	✓
Payback period analysis	N/A	✓
Process Services and Management		
Delivery Method	Design-Bid-Build (IPD lite)	Design Build
Project Priority	Replicable Process	System Performance Demonstration
Technology Selection Criteria	State of Shelf	State of Art

*Relative to ASHRAE 90.1-2007

The distinction on primary performance outcomes are shaped by the criteria set at earlier phases. For example, the EEB Hub team defined the scope of system technology selection with the proven technology which is available and accessible in current industry conditions (State of Shelf), whereas the EIC team used cutting edge new technologies that will also be tried and tested (State of Art). However, both

teams set their different values and goals at the earliest phase available. Later, in the process of design decision generation is deployed as below by EEB Hub team. Additionally, we have developed a comparative process models by using BPMN technique. The "to be" model is prepared based on RMI's retrofit recommendations derived from best practices. Additional activities and deliverables are added to the early design part of "as is" EEB and EIC project maps while lacking activities and deliverables facilitated in related processes are omitted as in the Appendix.

References

C.J. Anumba, A.F. Cutting-Decelle, A.N. Baldwin, J. Dufau, M. M. Bouchlaghem. "Integration of Product and Process Models as a Keystone of Concurrent Engineering in Construction: the ProMICE Project." 2nd European Conference on Product and Process Modelling . Amor R., 1998.

Chen, Z., Clements-Croome, D., Hong, J., Li, H. and Xu, Q. (2006) 'A multi-criteria lifespan energy efficiency approach to intelligent building assessment', Energy and Buildings, vol 38, no 5, pp393–409.

Conroy, C. (2012, September 12). Personal Interview

Gholap, A. K. and Khan, J. A. (2007) 'Design and multi-objective optimization of heat exchangers for refrigerators', Applied Energy, vol 84, no 12, pp1226–1239.

Hommes, L.J. "The Evaluation of Business Process Modeling Technique." Ph.D. Thesis. Delft: Delft University of Technology, 2004.

Laden Aldin, Sergio de Cesare. "A Comparative Analysis of Business Process Modelling Techniques ." U.K. Academy for Information Systems (UKAIS). Oxford, UK, 2009.

M.D. Bostenaru Dan. "Multi-criteria Decision Model for Retrofitting Existing Buildings." Natural Hazards and Earth System Sciences, 2004: 485-499

Martin Owen, Jog Raj. www.omg.org/bpmn. 04 27, 2011. http://www.omg.org/bpmn/Documents/6AD5D16960.BPMN_and_BPM.pdf.

Martinaitis V., Rogoza, A. and Bikmaniene, I. (2004) 'Criterion to evaluate the "twofold benefit" of the renovation of buildings and their elements', Energy and Buildings, vol 36, no 1, pp3–8.

Pelin Gultekin, Sinem Mollaoglu-Korkmaz, David R. Riley, Robert M. Leicht. "Process Indicators to Track High Performance Green Building Project Performance." ASCE Journal of Management in Engineering Management for Considerations on Sustainability. In Review.

Pelin Gultekin, Chimay J. Anumba, Robert M. Leicht. "Towards An Integrated Process Model And Decision Support System For High Performance Green Retrofits." 2013 Architectural Engineering Institute Conference. In Press.

Rey, E. (2004) 'Office building retrofitting strategies: Multi-criteria approach of an architectural and technical issue', Energy and Buildings, vol 36, no 4, pp367–372.

Rosenfeld, Y. and Shohet, I. M. (1999) 'Decision support model for semi-automated selection of renovation alternatives', Automation in Construction, vol 8, no 4, pp503–510.

S. Austin, A. Baldwin, A. Newton. "A Data Flow Model to Plan and Manage the Building Design Process." Journal of Engineering Design, 1996: 3-25.

Santamouris, M. (2008). *Advances in Building Energy Research*. Sterling, VA: Earthscan

DiBartolo, S. (2012, August 8). Personal Interview.

Zhu, Y. (2006) 'Applying computer-based simulation energy auditing: A case study', Energy and Buildings, vol 38, no 5, pp421–428.

APPENDIX: "To be" and "As is" process models

RMI- Retrofit

EEB HUB- Retrofit

EIC- Retrofit
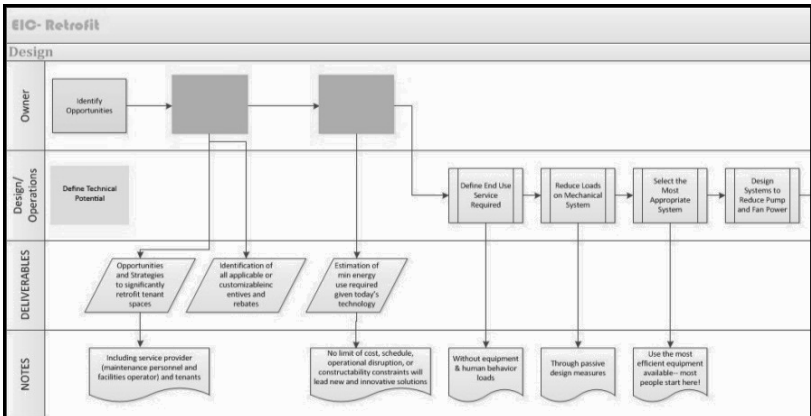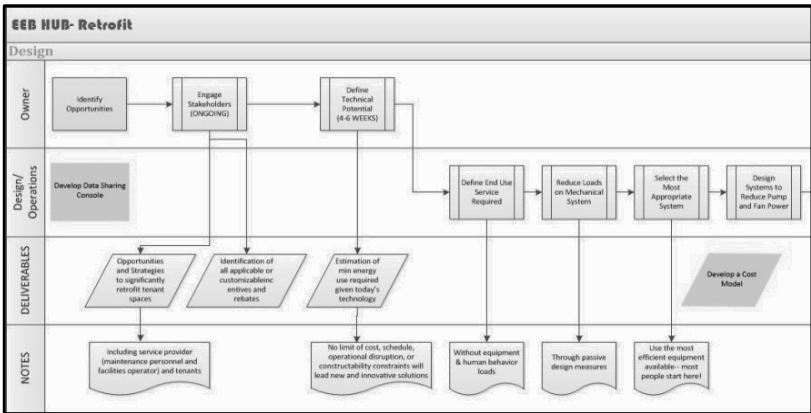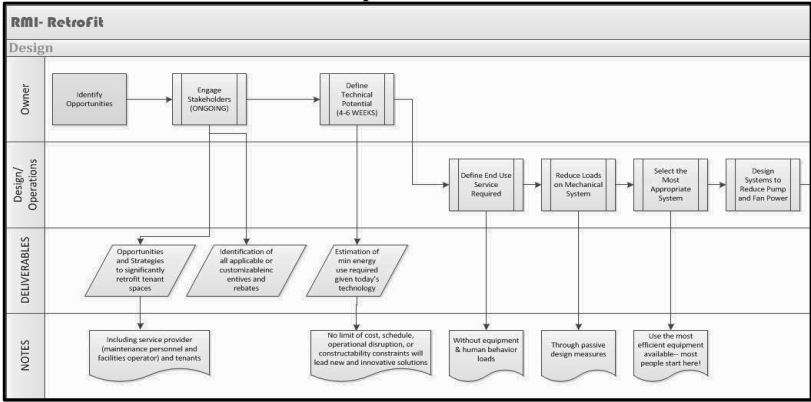
Robust Sustainability Management and Maintenance using Markov Decision Processes

Hadi Meidani[1] and Roger Ghanem[2]

[1] Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA 90089; PH: (323) 580-2641; emails: meidani@usc.edu
[2] Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, CA 90089; PH: (213) 740-9528; emails: ghanem@usc.edu

ABSTRACT

The quantitative sustainability assessment of urban systems relies on available predictive models for these complex systems of systems. Markov chains and Markov Decision Processes are among the mostly used tools to integrate the stochasticity in decisions related to infrastructure management. In the present paper, we investigate the effects of uncertainties in the characterization of these decision models and make the case for the development of rationales that can differentiate between them. We will explain the extension of the theory of random transition matrices developed by the authors to Markov Decision Processes and illustrate, using a numerical example, that the policies obtained by solving the deterministic transitions is not necessarily robust to the potential variabilities of transition rates. The resulting probabilistic framework for Markov Decision Processes will also enhance the maintenance of urban systems by rigorously quantifying the confidence in the sustainability assessment metrics.

INTRODUCTION

Markov Decision Processes (MDP) are efficient tools for decision-making [Puterman (1994)]. MDPs are particularly appealing for complex systems since they merely require a Markovian assumption. In modeling these systems, first principles may not be available or they may be too complicated to be amenable to computational analysis. Thus the choice of a Markov model yields simplicity both in interpretation and implementation. This paper describes finite state discrete time Markov chain models. The transition probabilities parameterize this model, which collectively constitute the transition matrix.

A MDP is completely specified by its transition matrix and reward or cost function. This function associate reward or cost associated with each decision option (also called action). Reward functions are commonly assumed to be also functions of the state of the system. Even though the MDPs are probabilistic models, minor variations in their parameters may lead to different performance. Thus, decisions made based on MDP with inaccurate parameters may result in suboptimal outcome [Iyengar (2005), El Ghaoui and Nilim (2005)]. It is, therefore, of paramount importance to accurately estimate these parameters and understand the errors associated with these estimates. Inevitable inaccuracies due to insufficient observation resources or noisy measurements hinder such estimation. Moreover, systems may be inherently uncertain. Their transition behavior may vary from one time to another and the Markovian assumption may be unrealistic for certain systems. Thus, it may be too simplistic, perhaps critically unreliable, too, to represent the parameters of an MDP as deterministic variables.

Other efforts have been exercised in addressing the role of parametric uncertainties on the decisions supported by the MDP. Authors in [El Ghaoui and Nilim (2005)] introduced approximate solution algorithms for the robust MDP with uncertain transitions. A percentile optimization approach was introduced in [Delage and Mannor (2010), Xu and Mannor (2009)] where a guarantee at some percentile level less than 100% was enforced for the optimality condition. Another approach was introduced in [Givan et al. (2000)] where bounded intervals were used for transition probabilities and reward functions. In this paper we construct probabilistic models for transition matrices that are consistent with known constraints and a Maximum entropy principle [Jaynes (1957)]. It is developed based on the theoretical foundation of Maximum Entropy random transition matrices proposed in [Meidani and Ghanem (2012)].

TECHNICAL BACKGROUND

In this section, we present the mathematical setting for the nominal MDP, based on which the stochastic variation of the following section is derived. We will consider the infinite horizon MDP. Let $\mathcal{S}$ denote the finite state space for the Markov model of the underlying system with cardinality $|\mathcal{S}|$, and $\mathcal{A}$ the finite action space with cardinality $|\mathcal{A}|$. Let $P_a(s, s')$ denote the probability that system in state s under action a will reside in state s' at the next time step, i.e. $P_a(s, s') = \Pr(s_{t+1} = s' \mid s_t = s, a_t = a) \ \forall t$. Let $R_a(s, s')$ denote the reward (or cost) received (or incurred) by taking the action a and the transition from s to s'. The reward (or cost) function could be assumed to only depend on the action and current state.

Given this mathematical model for the dynamical system, an optimal policy for a decision maker is to be found. A policy $\pi : \mathcal{S} \to \mathcal{A}$ is a function indicating which action the decision maker should take when the system is in different states, and is denoted by $a_t = \pi(s_t)$ for state s_t and action a_t. Our objective is to maximize total reward (or minimize total cost) over infinite time horizon, i.e. $\max_\pi \sum_{t=0}^\infty \alpha^t R_{a^t}(s_t, s_{t+1})$, where α is the discount factor which satisfies $0 \le \alpha < 1$, so that the total reward (or cost) is bounded. A dynamic programming approach, called value iteration, is used for finding the optimal policy, which is based on the iteration of the Bellman equation [Bellman

(1957)], according to

$$V^{k+1}(s) \leftarrow \max_a \sum_{s'} \left\{ P_a(s, s') \left(R_a(s, s') + \alpha V^k(s') \right) \right\} . \tag{1}$$

until convergence is achieved for every state s.

MARKOV DECISION PROCESSES WITH UNCERTAIN TRANSITIONS

Decisions made based on inaccurate MDP parameters may result in underestimation or overestimation of budget allocation. Both the reward functions and transition probabilities of the MDP can be modeled as uncertain quantities [Delage and Mannor (2010)]. Our current focus, however, is on transition uncertainties [Ghanem and Meidani (2012)]. Let $\mathbb{P}_a(\omega)$ denote the transition matrix associated with action a, i.e. $[\mathbb{P}_a(\omega)]_{ij} = P_a(i, j)$. Our objective in the present study is to investigate the implications of the uncertainty treatment of transition matrices on the decisions made based on the associated MDP. First, we will discuss different probability measures for the random transition matrices (RTM), based on which quantitative study on the response of the MDP can be carried out. Throughout the present paper, we assume the rows of the random transition matrix are independent from one another. Thus, we will describe the probability measure for a particular row of the random transition matrix denoted by $\boldsymbol{r}(\omega) = [r_1(\omega), \cdots, r_n(\omega)]$.

Dirichlet Distribution for Uncertain Transitions According to the Dirichlet distribution [Ferguson (1974)], the joint probability distribution of the multivariate $\boldsymbol{r}(\omega)$ is given by $f_{\boldsymbol{r}}^D(\boldsymbol{r}) = K \prod_{i=1}^n r_i^{\alpha_i - 1}$, where $\{\alpha_i\}$ are the parameters of the Dirichlet distribution and K is a normalization constant. The support of the Dirichlet distribution by definition is the following standard $n-1$-simplex $\{\boldsymbol{r} = [r_1(\omega), \cdots, r_n(\omega)] \mid \sum_{i=1}^n r_i = 1\}$. Due to its particular form of the support, this distribution has been considered for random transition matrices, which have the same normalization constraint, and also as a prior distribution on probability measures [Ferguson (1974)]. Dirichlet distribution is an appealing probability measure since it is conjugate prior to multinomial distributions, hence efficiency in its Bayesian update.

Maximum Entropy Distribution for Uncertain Transitions The probability measure for the multivariate $\boldsymbol{r}(\omega)$ can alternatively be estimated using the principle of Maximum Entropy (MaxEnt) [Jaynes (1957)]. The MaxEnt probability measure for the RTM was introduced in [Meidani and Ghanem (2012)]. It was assumed that the mean and standard deviations of the matrix components are available. On each row, one of the components can be written in terms of the other $n - 1$ components. Let $f_{r'}(r_1, \cdots, r_{n-1})$ denote this MaxEnt joint pdf for the $(n-1)$-dimensional random row $\boldsymbol{r}'(\omega)$ taking values in the following admissible set

$$V' = \left\{ \boldsymbol{r}' \subset \mathbb{R}^{n-1} \mid \sum_{i=1}^{n-1} r_i' \leq 1, \ \boldsymbol{r}' \geq 0 \right\} . \tag{2}$$

If only the mean values for transition matrices are available, we will have [Meidani and Ghanem (2012)]

$$f_{r'}^*(\boldsymbol{r}') = e^{(\mu-1)} \exp\left[\sum_{i=1}^{n-1} \lambda_i r_i\right] \mathbb{1}_{V'}(\boldsymbol{r}') \tag{3}$$

where $\mathbb{1}_{V'}$ is the indicator function for V'. If standard deviations are also included, the MaxEnt solution will instead have the following form

$$f_{r'}^*(\boldsymbol{r}') = e^{(\mu-1)} \exp\left[\sum_{i=1}^{n-1} \lambda_i r_i + \sum_{i=1}^{n-1} \eta_i r_i^2\right] \mathbb{1}_{V'}(\boldsymbol{r}') \, . \tag{4}$$

DYNAMIC PROGRAMMING UNDER TRANSITION UNCERTAINTY

Due to the uncertainty in the transition probabilities, a probabilistic treatment of the Bellman equation, i.e. Eq. 1, is necessary. In literature, the value under summation in Eq. 1 is denoted by Q, based on which the Q-learning algorithm is named. We have

$$Q(s,a) = \sum_{s'} P_a(s,s')(R_a(s,s') + \alpha V(s')). \tag{5}$$

At each step of the value iteration, for each state s, the action associated with the maximum Q value is selected.

In case of uncertainty transition probabilities, the Q values should be random, i.e.

$$Q(s,a)(\omega) = \sum_{s'} \mathbb{P}_a(\omega)(R_a(s,s') + \alpha V(s')). \tag{6}$$

Therefore, in order to determine the policy, one needs to predefine what the maximum Q value means for uncertain Q values . In what follows, we present different choice of basis for comparison between the random Q values, resulting in different solution algorithms for random MDP.

Three different approaches can be considered for the solution of the MDP with RTM. *Stochastic* optimization, suited to risk-neutral decision makers, the optimal policy is determined based on comparing the expectation of the Q value, i.e.

$$V^{k+1}(s) \leftarrow \max_a \mathbb{E}_{\mathbb{P}_a} \sum_{s'} \left\{ \mathbb{P}_a(\omega)\left(R_a(s,s') + \alpha V^k(s')\right)\right\} \tag{7}$$

Risk-averse decision makers care more about the risks associated with their decisions, rather than their expected consequences. In order to support the decisions of this group, *robust* or *worst-case* optimization approach has been introduced [El Ghaoui and Nilim (2005), Iyengar (2005)]. It involves in optimizing the policies in the event of worst case scenario using the following value iteration equation

$$V^{k+1}(s) \leftarrow \max_a \min_{\mathbb{P}_a} \sum_{s'} \left\{ \mathbb{P}_a(\omega)\left(R_a(s,s') + \alpha V^k(s')\right)\right\} \, . \tag{8}$$

An alternative approach is to consider the confidence intervals instead of the worst case scenarios. Worst case scenarios fail to provide quantitative risks associated with them. It would be more valuable if the risk averse decision makers are aware of how probable the scenarios are. A risk adjusted discounted performance of the optimization problem can serve this purpose. Formally speaking, the following value iteration scheme, called *chance-constrained* optimization, can furnish the decision makers with such information

$$V^{k+1}(s) \quad \leftarrow \quad \max_a \ y$$
$$\text{s.t.} \quad \Pr(\sum_{s'} \left\{ \mathbb{P}_a(\omega) \left(R_a(s, s') + \alpha V^k(s') \right) \right\} \geq y) \geq \eta. \tag{9}$$

where $\Pr(\cdot)$ denotes the probability of drawing a sample transition matrix from Dirichlet or MaxEnt distributions. By imposing this percentile constraint, we impose an η guarantee that the maximum random Q value is greater than its counterparts associated with other actions. It should be noted, that robust optimization is an especial case of percentile optimization for $\eta = 1$.

NUMERICAL DEMONSTRATION

In this section, we present the application of MDP with uncertain transitions to the problem of optimal infrastructure maintenance, in particular pavement maintenance. The deterioration process is modeled by a Markov chain. We used the data obtained from the highway pavements reported in [Carnahan et al. (1987)]. The Pavement Condition Index (PCI) was used to indicate the damage condition of the pavement. Eight different damage states were considered for the pavement. Three repair options (actions) were considered for the pavement: routine maintenance, two-inch overlay, and reconstruction. Under these three actions, the deterioration show different dynamics. The three transition matrices associated with the three actions, reported in [Madanat and Ben-Akiva (1994), Ghanem and Meidani (2012)], were used.

The implications of the uncertainty representation of the transition matrices on the decision variables were discussed in [Ghanem and Meidani (2012)], where different estimates of the cost were obtained, even though the policies remained unchanged. In this work, the costs pertaining to each action option were modified (see Table 1), so that the optimal policies are also sensitive to the optimization approach. A modified discount factor of 0.845 was used.

Let us consider a joint probability distribution for each (independent) row of the transition matrix in order to account for our ignorance. We assume the deterministic transition rates are the mean values. We also assumed a constant coefficient of variation equal to 30% for all the transition probabilities. In Fig. 1(a), the resulting marginal pdf of p_{12} is compared against their Dirichlet counterpart, obtained with the same observation data.

Given the tabulated nominal values for transition probabilities and costs, a nominal MDP can be solved to obtain the optimal policy. The optimal policy and its associated

(a)

(b)
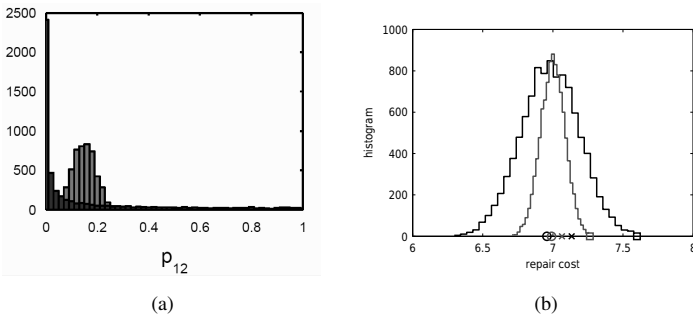
Figure 1: The pdf's and joint pdf's corresponding to the independent random transition probabilities outgoing states 1 to 5, plotted in figures (a) to (e), respectively.

costs are reported in Table 2. In order to investigate the impact of the transition uncertainties on the decisions, samples from the two distributions for the RTM were drawn, based on which the optimal policy was calculated using different approaches discussed in previous section. The MaxEnt RTM was constructed based on the mean values and a constant coefficient of variation of 30% for all the transition probabilities. An 80% percentile level was considered for solving the chance-constrained problem. Table 2 tabulates the optimal policies and their associated cost estimates. It can be seen the optimal policies are sensitive to the type of the optimization approach. Fig. 1(b) plot the predicted random cost associated with the optimal policy when transition matrices are considered to be random. Specifically, the costs associated with the routine repair (action 1) and overlay repair (action 2) are shown when the structure is in damage state 3. The markers in the figures refer to different choices of the quantity of interest. The three choices are shown; (1) the mean cost , (2) the cost at 80% percentile level and (3) the maximum cost. It can be seen that the way the quantities of interest compare are different. These figures can serve as an explanation for the difference observed in Table 2 in the optimal policies associated with stochastic, percentile, and robust MDP.

Table 1: Cost of different repair actions at different damage states Madanat and Ben-Akiva (1994)

state	routine	overlay	reconstruction
1	0.04	3.81	25.97
2	0.15	3.91	25.97
3	0.31	4.5	25.97
4	0.65	6.64	25.97
5	0.83	9.06	25.97
6	1.40	10.69	25.97
7	2.00	12.31	25.97
8	6.90	21.81	25.97

Table 2: The optimal policy and the respective total costs obtained by different versions of the MDP using 10,000 samples from Dirichlet and MaxEnt distributions. Asterisks indicate a policy different from the nominal MDP policy.

Type of MDP	state							
	1	2	3	4	5	6	7	8
policy (nominal)	1	1	1	2	2	2	2	3
cost (nominal)	2.05	4.25	6.48	8.65	12.93	16.44	20.16	27.97
policy (stochastic-Dir)	1	1	1	2	2	2	2	3
cost (stochastic-Dir)	1.94	4.23	6.42	8.56	12.88	16.37	20.11	27.90
policy (stochastic-MaxEnt)	1	1	1	2	2	2	2	3
cost (stochastic-MaxEnt)	2.04	4.27	6.48	8.65	12.95	16.44	20.16	27.97
policy (robust-Dir)	1	1	1	2	2	2	2	3
cost (robust-Dir)	11.01	12.98	15.19	17.61	21.89	25.56	30.80	36.94
policy* (robust-MaxEnt)	1	1	2	2	2	2	2	3
cost (robust-MaxEnt)	4.14	6.61	8.60	10.74	15.13	18.48	22.46	30.07
policy* (percentile-Dir)	1	1	2	2	2	2	2	3
cost (percentile-Dir)	3.77	6.68	8.35	10.45	15.07	18.21	22.37	29.73
policy* (percentile-MaxEnt)	1	1	2	2	2	2	2	3
cost (percentile-MaxEnt)	2.60	4.92	7.05	9.20	13.55	16.98	20.78	28.52

It should also be noted that the cost estimates based on the MaxEnt probability measure is considerably lower than that obtained by the Dirichlet distribution. That is because the probability mass in the MaxEnt measure is more concentrated and in the Dirichlet distribution is more spread.

CONCLUSION

A Markov Decision Process with uncertain transition property was studied. The MaxEnt model for the transition matrices was compared with the standard Dirichlet distribution. The implications of such uncertainty representation on the decision making procedure was demonstrated in an optimal maintenance problem for freeway pavement. It was shown that a different optimal policy may be obtained by considering the uncertainty in the transition rates.

References

Bellman (1957). R. Bellman. *Dynamic programming*. Princeton University Press, 1957.

Carnahan et al. (1987). JV Carnahan, WJ Davis, MY Shahin, PL Keane, and MI Wu. Optimal maintenance decisions for pavement management. *Journal of Transportation Engineering*, 113(5):554–572, 1987.

Delage and Mannor (2010). E. Delage and S. Mannor. Percentile optimization for markov decision processes with parameter uncertainty. *Operations Research*, 58(1):203–213, 2010.

El Ghaoui and Nilim (2005). L. El Ghaoui and A. Nilim. Robust solutions to markov decision problems with uncertain transition matrices. *Operations Research*, 53(5), 2005.

Ferguson (1974). T.S. Ferguson. Prior distributions on spaces of probability measures. *The Annals of Statistics*, pages 615–629, 1974.

Ghanem and Meidani (2012). R. Ghanem and Hadi Meidani. Random markov decision processes for sustainable infrastructure systems. In *Proceedings of the 3rd International Symposium on Advances in Urban Safety (SAUS2012), Nanjing, China.*, 2012.

Givan et al. (2000). R. Givan, S. Leach, and T. Dean. Bounded-parameter markov decision processes. *Artificial Intelligence*, 122(1):71–109, 2000.

Iyengar (2005). G.N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, 30(2):257–280, 2005.

Jaynes (1957). E.T. Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.

Madanat and Ben-Akiva (1994). S. Madanat and M. Ben-Akiva. Optimal inspection and repair policies for infrastructure facilities. *Transportation Science*, 28(1):55–62, 1994.

Meidani and Ghanem (2012). H. Meidani and R. Ghanem. Uncertainty quantification for markov chain models. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, pages 377 –382, Dec 2012. doi: http://dx.doi.org/10.1063/1.4757645.

Puterman (1994). M.L. Puterman. *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, Inc., 1994.

Xu and Mannor (2009). H. Xu and S. Mannor. Parametric regret in uncertain markov decision processes. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 3606–3613. IEEE, 2009.

Influence of Social Sub-networks on Energy Conservation from Occupancy Interventions in a Typical US Commercial Building

Elie Azar, S.M. ASCE[1] and Carol C. Menassa, PhD, A.M. ASCE[2]

[1]PhD Student, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 2256 Engineering Hall 1415 Engineering Drive, Madison, WI 53706; PH (608) 262-3542; FAX (262) 911-5199; email: eazar@wisc.edu
[2]M.A. Mortenson Company Assistant Professor, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 2318 Engineering Hall 1415 Engineering Drive, Madison, WI 53706; PH (608) 890-3276; FAX (262) 911-5199; email: menassa@wisc.edu

ABSTRACT

Occupancy focused interventions used to reduce building energy use (e.g., feedback) are exhibiting low and un-sustained energy reductions. Research on residential buildings has shown that the type of building social network determines the effectiveness of these interventions. However, the findings cannot be directly applied to commercial buildings given their more complex social structures formed by independent entities (i.e., companies) with different organizational structures and cultures. Therefore, the objective of this study is to evaluate the influence of social sub-networks on the energy conservation from occupancy focused interventions. An agent-based model was developed for this purpose simulating sub-networks in a typical commercial building environment. Occupancy interventions were then tested while varying the number of sub-networks in the building. Statistical analyses were finally performed showing that modeling sub-networks significantly affects the effectiveness of occupancy interventions and need to be accounted for in the design of more efficient interventions for commercial buildings.

INTRODUCTION

In the United States (US), 19 percent of total energy use is consumed by commercial buildings. The energy demand of this sector is also growing at a rate of 2.9 percent per year which is higher than the rate of any other sector of the economy (EIA 2011). There are two different approaches through which energy conservation can be achieved. The first one is the technological approach, which includes installing more energy efficient technologies (EPA 2010), while the second approach aims at conserving energy by helping occupants adopt more conservative energy use characteristics (Carrico and Riemer 2011; Abrahamse et al. 2005). The focus of this paper is on the latter approach where different types of interventions are known to have helped building occupants change their energy use characteristics and save energy. These interventions include: (1) peer pressure (e.g., Allcott and Mullainathan

2010), (2) feedback techniques (e.g., Peschiera et al. 2010), (3) energy training and education (e.g., Verplanken and Wood 2006), and (4) green social marketing campaigns (e.g., McKenzie-Mohr 2000). While the energy saving potential of the mentioned techniques is significant, studies have rarely been conducted in the commercial building sector. Furthermore, the promoted conservative energy use characteristics were rarely maintained over time and relapses to pre-intervention energy use levels were commonly observed (Abrahamse et al. 2005; Jackson 2005).

In an effort to address the above mentioned limitations, recent research has shown that the efficiency of interventions (e.g., feedback) is highly dependent on the social network's characteristics of the building under study (Anderson et al. 2012; Chen et al. 2012). In these studies, one building was typically modeled as one large social network. This assumption is realistic in residential buildings where occupants are randomly distributed (e.g., students in a dormitory building) without a significant level of clustering (Chen et al. 2012; Peschiera et al. 2010).

However, the obtained results cannot be applied in commercial buildings, which typically have more complex social structures due to the possible presence of multiple independent entities within the same building (i.e., companies) (EIA 2003). As shown by Groeber et al. (2009), companies typically have different cultural, organizational, and social structures, which affect how people interact and respond to interventions (Mason et al. 2007). In fact, the occupants clustered in the social sub-networks created by companies interact regularly and have higher chances to be influenced by their peers rather than by occupants from another entity (Apolloni and Gargiulo 2011; Nolan et al. 2008). Such dynamics of interaction in commercial buildings need to be accounted for to truly understand the impact of interventions on individual (i.e., occupant) and system (i.e., building) levels (Mason et al. 2007).

The objective of this study is to model a typical commercial building with its social sub-networks and determine whether using the sub-network approach, rather than the traditional one network approach, impacts the effectiveness of interventions. Results can be used to shape future occupancy interventions for more effective and sustained energy reductions in commercial buildings. Prior to proceeding with the next sections, the authors would like to note that some of the results shown in this paper are also part of a paper submitted to the ASCE Journal of Computing in Civil Engineering, Special Issue on Computational Approaches to Understand and Reduce Energy Consumption in the Built Environment (Azar and Menassa 2013).

METHODOLOGY

A three-step methodology was developed to achieve the study's objectives: (1) model development, (2) parametric variation, and (3) statistical analysis of results.

General Model Description

The model used in this research was developed using the agent-based modeling (ABM) platform from *Anylogic,* a widely used simulation software. ABM is a computational method that enables researchers to create and experiment with models composed of agents interacting within an environment. It is considered as one of the most appropriate techniques for modeling human behavior and interactions, social networks, and opinion dynamics (Gilbert 2008; Edmonds et al. 2007).

The main purpose of the proposed model was to emulate a commercial building environment with different social entities (i.e., sub-networks) to test how occupancy interventions and interactions impact the energy consumption of the building. The agents represent the building occupants who are characterized by two state variables: (1) *Energy Intensity*, which defines the energy consumption level of the agent, and (2) the energy intensity variability, referred to in this paper as *Variability*, which defines how the variable energy use behavior of an agent affects his/her chances of adopting new energy use characteristics. As an example, an occupant with an energy intensity of 20 kWh/m2/person/year and a *Variability* of 9 kWh/m2/person/year can only be influenced by peers whose energy intensities do not differ by more than 9 kWh/m2/person/year from his/her own *Energy Intensity* (i.e., 20 ± 9 kWh/m2/person/year). This concept was adopted from the "Relative Agreement" RA model of Deffuant et al. (2002) as detailed in the following section.

Agents in the model are grouped in different sub-networks representing the different social entities (i.e., companies) in the building. Each entity can be assigned a network type distinct from the other entities (e.g., small word, scale-free, random), which defines the connections and interactions between the agents of a single entity. The energy use characteristics of occupants can change due to (1) peer influence (based on the RA principle), and/or (2) due to an occupancy intervention set by the modeler to emulate an occupancy focused interventions such as energy education or feedback. This in turn might motivate additional energy conservation through peer influence as the occupants in the same sub-network further interact with each other.

Relative Agreement Principle

As defined by Deffuant et al. (2002), the RA principle or model defines the dynamics of peer-influence between people with different 'opinions'. The RA model has been widely used in literature and was adapted to several fields such as the study of fashion trends, religious extremism, and political opinions (Gilbert 2008; Deffuant et al. 2002). In this study, the RA model was adapted to emulate changes in the energy use characteristics of building occupants. The main principles of the RA model are presented below with analogies made to illustrate and justify its applicability to the proposed study of energy use behaviors.

Principle 1: Each agent in the RA model is characterized by a continuous 'opinion' variable that can change over time. Similarly, each building occupant is characterized by a continuous 'Energy Intensity' variable that can also change over time. Studies in literature confirm that occupancy energy use characteristics can in fact change from interventions, affecting how much energy each occupant consumes (Peschiera et al. 2010).

Principle 2: Each agent is characterized by an 'uncertainty' variable which represents his/her openness to different opinions. Each building occupant is characterized by a 'Variability' variable which represents his/her openness to different energy use characteristics. For instance, an occupant with strict energy consumption patterns (e.g., strong habits of leaving lights on) would be considered to have a low Variability and would be harder to influence than a person with a high variability (e.g., flexible energy use patterns) (Verplanken and Wood 2006).

Principle 3: The RA model assumes that people adapt their behaviors or opinions to conform to the behavior of their peers. The same social influence applies

to the study of pro-environmental behaviors such as energy conservation (Gockeritz et al. 2010). For instance, if occupants A, B, and C leave the lights on afterhours 5, 1, and 4 days/week respectively, an influence of B on A is expected to be higher than the one of C on A since C and A have relatively similar energy use characteristics.

Principle 4: The RA model assumes that agents not only influence each others' 'opinions' but also their 'uncertainties'. Using the same example as above, the influence of B on A might also lead to an increase in the Variability of the energy use characteristics of A. For instance, if A initially had a relatively low variability (e.g., 5 ± 1 day/week), peer-influence might increase this variability (e.g., 5 ± 3 days/week), making this occupant more open to new energy use characteristics.

Model subsections and execution

The main sub-models that define the model's execution steps are shown in Figure 1 and are detailed in the following paragraphs.

Sub-model 1: Called at time 0, the 'Agents and Sub-networks Creation' sub-model creates the agents and assigns them to different sub-networks. Then, it generates the connections between agents for each sub-network based on the topology specified by the user. In this paper, data was gathered from the Commercial Building Energy Consumption Survey (CBECS) to determine the characteristics of a typical US medium-size commercial building (EIA 2003). The collected data showed that on average, a typical building is occupied by 3 companies with 11 employees per company. Therefore, the model was initialized with 3 sub-networks and 11 agents per sub-network. The level of granularity obtained from the CBECS data was made possible due to a previous work by the authors combining the 20 published CBECS data files (Azar and Menassa 2012).

Sub-model 2: Next, the 'Energy Use Characteristics Initialization' sub-model initializes the *Energy Intensity* and the *Variability* variables specific to each occupant. Here again, CBECS data was gathered on the energy use characteristics of medium-size commercial buildings' occupants expressed in kWh/m2/person/year. Since energy use data is always positive with values falling in between two extreme levels, the collected data was compared to known restricted distributions and a Log-normal distribution fit was obtained with a Kolmogorov Smirnov p-value of 0.036, confirming the goodness of fit. The obtained parameters of the Log-normal distribution (μ=1.626, σ=0.875, and γ=-0.099) were then used to initialize the *Energy Intensity* of each building occupant as shown in Eq. 1.

As for the *Variability*, it was initialized based on the concepts of extremism in opinions and 'Bounded Confidence', which were used in the RA model of Deffuant et al. (2002). This resulted in low initial *Variability* values for occupants with extreme *Energy Intensities* (i.e., very low or very high energy consumers), and high values for moderate energy users in order to represent their higher acceptance of different behaviors. Additional details can be found in Azar and Menassa (2013).

$$Energy\ Intensity_{Occupant\ i} = lognormal\ (\mu, \sigma, \gamma) \tag{1}$$

Sub-model 3: The 'Peer Influence' sub-model is launched at each time step, simulating the potential change in occupants' characteristics following their contact with a peer. During one time step, each agent sends a message to a random connected

agent in the same sub-network to change behavior. When received, the model stores the *Energy Intensity* and *Variability* values for both agents and applies formulas based on the RA principle (Deffuant et al. 2002). In summary, let: i be the index of the agent sending the message, j the index of the receiving agent, *EI* the value of the *Energy Intensity* variable, and *Var* the *Variability* in energy intensity. An 'overlap' is first calculated to determine how close the two behaviors of agents are (See Eq. 2). Then, if the inequality of Eq. 3 is true, the energy use behavior of the receiving agent changes and both *EI* and *Var* are updated according to Eq. 4 and Eq. 5 respectively. In these equations, σ is a constant controlling the speed of the dynamics and does not represent an actual physical parameter (Deffuant et al. 2002). A default value of 0.5 was therefore used avoiding extremely slow or fast dynamics (i.e., σ close to 0 or 1).

$$Overlap_{ij} = \min(EI_i + Var_i, \ EI_j + Var_j) - \max(EI_i - Var_i, \ EI_j - Var) \qquad (2)$$

$$Overlap_{ij} > Var_i \qquad (3)$$

$$EI_j = EI_j + \sigma\left(\frac{Overlap_{ij}}{Var_i}\right) \times (EI_i - EI_j) \qquad (4)$$

$$Var_j = Var_j + \sigma\left(\frac{Overlap_{ij}}{Var_i}\right) \times (Var_i - Var_j) \qquad (5)$$

Sub-model 4: The 'Discrete intervention' sub-model checks if an occupancy intervention was scheduled at the current time-step. The time of occurrence and the energy reduction effectiveness of the event are determined at time 0 by the modeler. When an intervention is scheduled, the model reduces the *Energy Intensity* of occupants in a stochastic manner while maintaining the average effectiveness specified by the user (See Eq. 6). In this paper, the intervention was scheduled to occur at time 0 with an average efficiency of 10 percent, which was based on previous studies in literature that quantified the efficiency of typical interventions between 3 and 17 percent (Carrico and Riemer 2011; Pickens et al. 2002).

$$EI_i = EI_i \times [1 - Uniform\ (0, Average\ Intervention\ Efficiency \times 2)] \qquad (6)$$

Sub-model 5: Called at each time step, the 'Building Energy Calculation' sub-model first stores the *Energy Intensity* and *Variability* variables of all building occupants and calculates the total building energy use.

Sub-model 6: 'Convergence Check' occurs at the end of each time step where the model checks if the change in total building energy use over the last 50 iterations falls below 1 Watt-Hour (0.001 kWh). When convergence occurs, the simulation ends and the model exports the results to text files for analysis. Otherwise, the model moves to the next time step and goes back to sub-model 3 as shown in Figure 1.
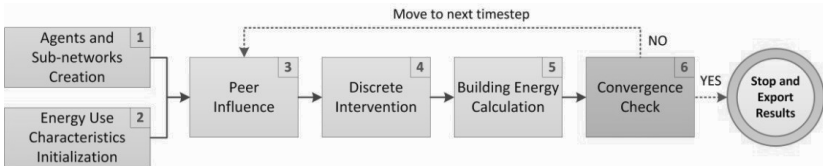
Figure 1. Sub-models' execution (Adapted from Azar and Menassa 2013)

Parametric Variation and Statistical Analyses

Following the initialization of the model, a parametric variation was performed on the number of sub-networks to test whether accounting for sub-networks impacts the effectiveness of interventions. This was essential to justify the need to model the social structure of commercial buildings using sub-networks. The analysis was repeated for the two most commonly encountered social network topologies in buildings: scale-free and small world (Mason et al. 2007; Watts and Strogatz 1998). It is important to mention that no connections between sub-networks were considered in order to fairly compare the traditional approach (i.e., one network) to the proposed one (i.e., multiple sub-networks).

In summary, six different scenarios were generated consisting of models with 1, 3, and 5 networks, in addition to small world and scale free topologies. Each scenario was run 1000 times and the values of total building energy use (TE) and convergence time (CT) were collected. Therefore, in order to compare the TE and CT of two scenarios, it was essential to perform statistical analyses on the means of the 1000 runs of each scenario.

The results of the six scenarios are shown in the upper graphs of Figure 2 highlighting the obtained means and 95 percent confidence intervals of TE and CT. Given that the main objective of this research was to determine if using sub-networks rather than traditional one network approach impacts the effectiveness of interventions, two sample tests were also performed comparing the results of scenarios with multiple sub-networks (Scenarios 2, 3, 5, and 6) to the single network ones (Scenarios 1 and 4). As shown in the bottom part of Figure 2, the differences in means were different than 0 with p-values lower than 0.05 for TE as well as for CT. Hence, the results prove that modeling sub-networks significantly impacts the effectiveness of energy saving interventions.
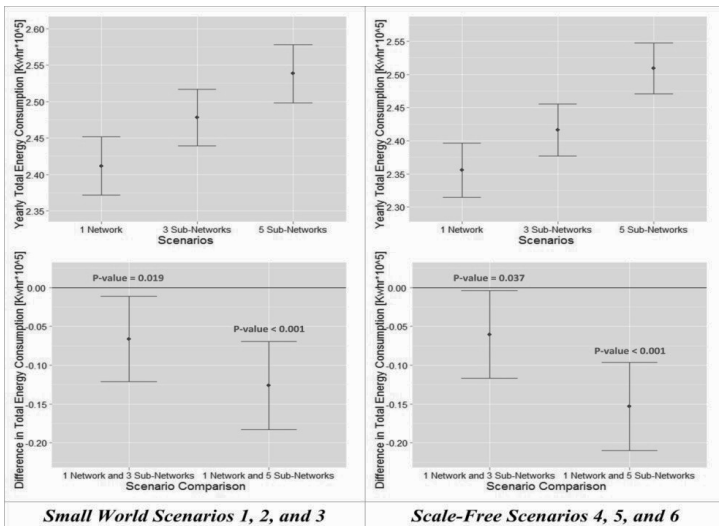
Figure 2. Total Energy Consumption Results

CONCLUSION

This paper presented an agent-based modeling framework that accounts for the social sub-networks present in a typical US commercial building. Occupancy focused interventions in addition to peer interactions and influence were modeled, emulating the different factors that can reduce building energy use through improved occupancy energy use characteristics. Parametric variations were then performed on the model proving that the traditional single-network modeling is not applicable to commercial buildings. This confirmed the need for the proposed approach that accounts for the different social entities that typically occupy a commercial building (i.e., companies) using sub-networks. As part of future research, the authors will monitor and collect data from a large number of buildings to test interventions and confirm the predictive validity of the model. The proposed framework can then be developed as a testing and optimizing tool for occupancy interventions, to be used prior to implementing these interventions in actual buildings. This is expected to help design more efficient and sustainable occupancy intervention strategies, significantly reducing energy consumption and carbon emissions in commercial buildings.

ACKNOWLEDGEMENTS

REFERENCES

Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005). "A review of intervention studies aimed at household energy conservation." Journal of Environmental Psychology, 25(3), 273–291.

Allcott, H., and Mullainathan, S. (2010). "Behavior and energy policy." *Science*, 327(5970), 1204–1205.

Anderson, K., Lee, S., and Menassa, C. C. (2012). "Effect of social network type on building occupant energy use." *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings - BuildSys '12,* Toronto, ON, Canada.

Apolloni, A., and Gargiulo, F. (2011). "Diffusion Processes through Social Groups' Dynamics." Advances in Complex Systems, 14(2), 151–167.

Azar, E., and Menassa, C. C. (2013). "A framework to evaluate energy saving potential from occupancy interventions in typical US commercial buildings". *Journal of Computing in Civil Engineering, ASCE* (Submitted).

Azar, E., and Menassa, C. C. (2012). "A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings." *Energy and Buildings*, 55(1), 841–853.

Carrico, A. R., and Riemer, M. (2011). "Motivating energy conservation in the workplace: An evaluation of the use of group-level feedback and peer education." *Journal of Environmental Psychology*, 31(1), 1–13.

Chen, J., Taylor, J. E., and Wei, H. (2012). "Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation." *Energy and Buildings*, 47, 515–524.

Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. (2002). "How can extremism prevail? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulation*, 5(4), 1–26.

Edmonds, B., Hernández, C., and Troitzsch, K. G. (2007). "Social Simulation Technologies, Advances and New Discoveries". *Information Science Reference*, Hershey, PA.

Gilbert, G. N. (2008). "Agent-Based Models". *Sage Publications, Inc*, London, United-Kingdom.

Gockeritz, S., Schultz, P. W., Rendo, T., Goldstein, N J, and Griskevicius, V. (2010). "Descriptive normative beliefs and conservation behavior: The moderating roles of personal involvement and injunctive normative beliefs." 40(3), 514–523.

Groeber, P., Schweitzer, F., and Press, K. (2009). "The case of local cultures." 12(2), 1–22.

Jackson, T. (2005). "Motivating Sustainable Consumption: a Review of Evidence on Consumer Behaviour and Behavioural Change". *Centre for Environmental Strategy, University of Surrey*, Surrey, United-Kingdom.

Mason, W. a, Conrey, F. R., and Smith, E. R. (2007). "Situating social influence processes: dynamic, multidirectional flows of influence within social networks." *Personality and social psychology review*, 11(3), 279–300.

McKenzie-Mohr, D. (2000). "Promoting Sustainable Behaviour: An Introduction to Community-Based Social Marketing." *Journal of Social Issues*, 56(3), 543–554.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, Noah J, and Griskevicius, V. (2008). "Normative social influence is underdetected." *Personality & social psychology bulletin*, 34(7), 913–23.

Peschiera, G., Taylor, J. E., and Siegel, J. A. (2010). "Response-Relapse Patterns of Building Occupant Electricity Consumption following Exposure to Personal, Contextualized and Occupant Peer Network Utilization Data." *Energy and Buildings*, 42(8), 1329–1336.

Pickens, P. M. (2002). "Community-Based Social Marketing as a Planning Tool: Community and Regional Planning". *Masters Project, University of Oregon, University of Oregon-Architecture and Allied Arts Department*, Eugene, OR.

US Energy Information Administration (EIA). (2011). "2011 Annual Energy Outlook". *EIA*, Washington, DC.

US Energy Information Administration (EIA). (2003). "Commercial Buildings Energy Consumption Survey." <http://www.eia.gov/consumption/commercial/> (Mar. 1, 2013).

US Environmental Protection Agency (EPA). (2010). "ENERGY STAR ® and Other Climate Protection Partnerships". *2010 Annual Report*, Washington, DC.

Verplanken, B., and Wood, W. (2006). "Interventions to Break and Create Consumer Habits". *Journal of Public Policy & Marketing*, 25(1), 90–103.

Watts, D. J., and Strogatz, S. H. (1998). "Collective dynamics of 'small-world' networks." *Nature*, 393(6684), 440–2.

Coupling Distributed Energy Simulation and Occupancy Models for Comprehensive Building Energy Consumption Analysis

Carol C. Menassa[1], Vineet R. Kamat[2], SangHyun Lee[3], Elie Azar[4], Chen Feng[5], Kyle Anderson[6]

[1]M.A. Mortenson Company Assistant Professor, Department of Civil and Environmental Engineering, University of Wisconsin-Madison; email: menassa@wisc.edu

[2]Associate Professor, Department of Civil and Environmental Engineering, University of Michigan; email: vkamat@umich.edu

[3]Assistant Professor, Department of Civil and Environmental Engineering, University of Michigan; email: shdpm@umich.edu

[4]PhD Student, Department of Civil and Environmental Engineering, University of Wisconsin-Madison; email: eazar@wisc.edu

[5]PhD Student, Department of Civil and Environmental Engineering, University of Michigan; email: cforrest@umich.edu

[6]PhD Student, Department of Civil and Environmental Engineering, University of Michigan; email: kyleand@umich.edu

ABSTRACT

We propose a conceptual framework that couples energy modeling with occupancy characteristics and energy use data to achieve comprehensive building energy consumption analysis. Specifically, we aim: 1) to couple distinct and spatially distributed simulation models and synchronize their data exchange; and 2) to demonstrate the coupled simulation through a hypothetical case example of a building.

This framework has been developed based on the principles defined in the High-Level Architecture (HLA) that enables distributed computing. In other words, the HLA-compliant federation allows federates (e.g., models) to communicate with each other and exchange relevant information to achieve the global objective of analyzing and reducing the building's energy use. A case study of a typical commercial building illustrates how the federates coordinate data synchronization and run in a distributed fashion. This example tests feedback frequency to building occupants on the building's energy use and illustrates the potential application of the framework to study energy interventions in buildings.

INTRODUCTION

Significant advances have been made in energy modeling, building monitoring and sensing devices aiming to reduce energy consumptions in buildings. For example, advances in energy simulation software (e.g., Energy-10, eQuest, DOE2 and EnergyPlus) allow architects and engineers to develop significantly detailed models that predict energy consumption of a building. On the other hand, the use of Energy Management Control Systems (EMCS), which consists of both hardware and software components to control and monitor building operations, allows facility managers to perform activities such as reducing demand charges by managing and scheduling equipment loads (EIA 2012; Andrews and Krogmann 2009). Another approach to ensure energy efficiency during operations phase focuses on occupancy education and feedback in residential buildings (Peschiera and Taylor 2012; Peschiera et al. 2010; Abrahamse et al. 2005) and commercial buildings (Azar and Menassa 2012). Whether it is energy management or occupancy interventions, building stakeholders can significantly benefit from coupling these resources under one simulation model where they can test strategies and analyze their impact on the building before their actual implementation. Such a simulation environment can reduce the facility manager's uncertainty in choosing technical and behavioral intervention strategies. In this sense, the main limitations of the de-coupled approach to building design and operation are: (1) a single and monolithic model cannot simulate complex processes within a domain with the required fidelity and detail; (2) a single set of model developers cannot have expert knowledge in all the details of a domain to be simulated; and (3) those who are managing a building's energy systems generally do not have the training to utilize the multiple programs.

The premise of this paper is that the de-coupled approach to simulation, modeling and energy management prevents the benefits of existing tools used in the design phase (e.g., EnergyPlus) from being extended to the operation phase of the building. Particularly, these existing energy models are very dependent upon data input by the designer/engineer, which limits their ability to simulate complex processes that use the data obtained during building operation phase. Therefore, we propose an extensible simulation framework that allows for the coupling of distinct and spatially distributed energy and occupancy simulation models, and synchronization of their data exchange through the use of High-Level Architecture (HLA: IEEE 1516). An extended version of this paper has been submitted and is currently in review with the ASCE Journal of Computing in Civil Engineering, Special Issue on Computational Approaches to Understand and Reduce Energy Consumption in the Built Environment (Menassa et al. 2013).

OBJECTIVES

The objectives of this paper are to: (1) investigate the applicability of the HLA to provide a platform to develop an extensible coupling of building energy analysis and occupancy models; (2) describe the prototype of the HLA framework that was developed for this purpose; and (3) provide a proof of concept for the prototype model through an application to a case study.

COUPLED BUILDING ENERGY SIMULATION

Several research efforts have been made to couple different building systems for control and testing either through direct coupling of energy simulation programs or through the use of a modular middleware. Direct coupling work includes: heat and air flow analysis in buildings (Hensen 1999), integration of multiple geographically distributed simulation applications within a building design tool (Lam et al. 2002), building energy simulation and Computational Fluid Dynamics (CFD) for prediction of energy and indoor environment (Zhai and Chen 2005), and component system packages for HVAC design (Trecka et al. 2006). A major limitation of these models is that they have been developed to address a specific application. As a result, the potential of these models to be reused and built upon by others is reduced.

On the other hand, the Building Controls Virtual Test Bed (BCVTB) developed at Lawrence Berkley National Laboratory (Pang et al. 2012; Wetter 2011) uses Ptolemy II (Brooks et al. 2007) as a modular middleware to couple simulation programs. Limitations of BCVTB are related to Ptolemy II since it only allows the modeling, simulation, and design of concurrent, real-time and embedded systems without great flexibility.

MODEL DESCRIPTION

In order to achieve the objectives, we applied the principles defined in the HLA proposed by the US Department of Defense (DOD) (Kuhl et al. 1999) to couple energy simulation. The DOD HLA is a collection of general rules that manage the development of complex, interoperable simulations in a distributed network environment. The HLA guidelines have been standardized by IEEE (Institute of Electrical and Electronics Engineers), and specifically developed to enable scalability, extensibility, and interoperability. In addition, they allow for shared model development effort for complex multidisciplinary problems and the reuse and assembly of multiple (perhaps existing) models in different contexts as part of a larger, interdependent, complex simulation (Kuhl et al. 1999).

In any simulation framework that applies the HLA principles, the HLA rules must be enforced if a federate (i.e., single simulation model) or federation (i.e. collection of multiple running and interacting federates) is to be regarded as HLA compliant. The HLA interface specification defines the functional modes of interaction between multiple federates and the framework's Run-Time Infrastructure (RTI). The RTI is software that must conform to the HLA specifications and provide simulation facilitation services (e.g., coordinates the synchronization and transfer of data between federates). The Object Model Template (OMT) prescribes standards for defining HLA object modeling information, which includes the data to be handled by the RTI when a simulation federate executes.

The biggest motivation to apply the HLA principles is that no single simulation or model in any domain can satisfy all uses and users (Dahmann et al. 1998). However, with the HLA principles, different simulations models that have been developed or will be developed can be composed together and form a HLA-compliant federation. The important tasks in developing the HLA-compliant

federation include: (1) investigating the requirements for creating a HLA-compliant energy simulation framework, and (2) establishing the necessary computing infrastructure to couple different models (i.e., middleware). In this research, CERTI HLA was selected and customized for the development of this federation. CERTI is is open-sourced and has multiple language bindings (e.g. C++/JAVA/Matlab) whose flexibility is critical for scalability and extensibility.

　　　　One of the most important capabilities from HLA RTI is enabling seamless communication between processes. Each simulation model/process (i.e. federate) interacts locally with an RTI Ambassador process (RTIA) in a whole simulation system (i.e., federation). Specifically, the RTIA listens to both the federate and the RTI. When a RTIA receives a message, RTI delivers it to the other interested RTIAs. By enabling such communications throughout all the constituent federates (e.g., making them HLA-compliant), a larger coupled simulation system can be realized.

Design of Energy Simulation Federation: Figure 1 (Menassa et al. 2013) illustrates how two federates, a building energy prediction federate (DOE2 Federate) and a building occupancy federate (Anylogic Federate), exchange an energy consumption parameter and a behavior level parameter. The Anylogic Federate is a computational agent-based model that simulates occupancy as a variable element by assigning attributes and characteristics to building occupants such as energy consumption estimates that correspond to different and changing energy consumption behaviors of occupants over time (Azar and Menassa 2012). The DOE2 federate uses BIM (Building Information Modeling) to generate the initial DOE2 energy simulation federate input (Kim and Anderson 2012).
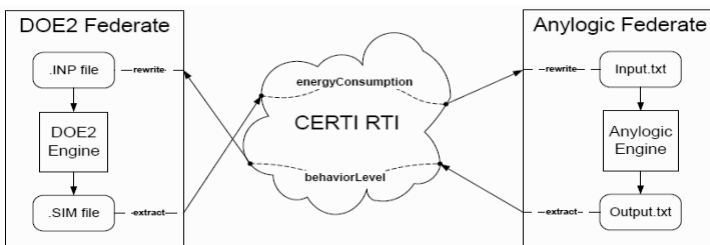
Figure 1: Data flow of energy simulation federation (Menassa et al. 2013)

　　　　Once this energy simulation is initiated, the DOE2 Federate will first read an initial input file (.INP file) generated by ifcXML (a neutral file format of BIM), in order to populate much of the information required for energy simulation in DOE2 from a building information model. The energy simulation federate invokes the DOE2 Engine from a command line to perform the initial energy estimate. Data from the estimate is then extracted from its output file (.SIM file) to compute the energy consumption parameter "energyConsumption" (see Figure 1). After it publishes this "energyConsumption" parameter through CERTI RTI, the DOE2 Federate will wait for CERTI RTI to update the behavior level parameter "behaviorLevel" from the Anylogic Federate for further simulation. When CERTI RTI receives the newly published "energyConsumption" parameter, it will immediately broadcast this value

to those federates who are interested in (i.e., subscribe to) the value. In this case, the Anylogic Federate that has been waiting for the "energyConsumption" parameter, will rewrite the newly received value to an "Input.txt" file. This file serves as the input of the Anylogic Engine (see Figure 1). Since the Anylogic engine is not open-sourced and can't be invoked from command line, the authors circumvent the problem by designing an Anylogic model which runs endlessly. Whenever it finds out the "Input.txt" file has been rewritten, the Anylogic model will start a new building occupancy simulation based on the new value reflected from this new "Input.txt", and writes its output, i.e. "behaviorLevel", to "Output.txt". As long as Anylogic Federate detects that "Output.txt" is rewritten, it will: 1) extract its value; 2) publish to CERTI RTI as the new value of "behaviorLevel"; and 3) wait for the next update of "energyConsumption". Now similarly, CERTI RTI will broadcast the newly updated "behaviorLevel" to its audience, DOE2 Federate that will rewrite the INP file accordingly for the next simulation iteration (Figure 1) (Menassa et al 2013).

In order to achieve the above described inter-process data exchange, a data exchange description file (.FED file) is designed based on what data is going to be sent through the CERTI RTI. When each federate tries to initialize itself, this FED file has to be provided so that the federate can either create a new federation or join an existing federation. The invoking sequence of the CERTI HLA API is thus divided into three phases: 1) preparation, 2) main simulation loop, and 3) clean up.

CASE STUDY

A case study was conducted to test the proposed framework. A generic one-story office building in Ann Arbor, MI was chosen. The building has 20 occupants and the baseline building systems schedules are set to conform to ASHRAE 90.1-2007 and thermal set points to ASHRAE 55-2010 (ASHRAE 2010; ASHRAE 2007). This building is being tested to determine how changes in building occupancy energy consumption patterns due to feedback intervention will affect the energy use in the building. It is assumed that the energy consumption estimates provided by DOE2 are representative of the actual energy consumption levels of the building. Feedback is an intervention technique that provides building occupants with their energy consumption levels, as well as target energy use levels.

Description of the Federate and data exchange through CERTI RTI: Modeling an intervention (e.g., feedback), which occurs during building operation, requires a certain level of coupling and communication between the energy simulation software (e.g., DOE-2) and the occupancy behavior software (e.g., an agent based model). In addition, ArchiCAD was used to model the building under study and import the 3D model in DOE-2 to simulate the building energy performance.

In the study of feedback, real-time energy levels are critical for a realistic modeling of the feedback concept. Thus, actual energy consumption levels generated by DOE-2 were imported to Anylogic and communicated to agents, who adapt their behavior accordingly. It was assumed that the 20 occupants of the building were initially evenly split between two categories: Green and non-Green. The influence to change behavior, caused by feedback, can occur in two ways: Green occupants can

convert their non-Green peers to the Green category and vice-versa. In the model, each category of occupants (i.e., Green and non-Green) has a Level of Influence (LI) parameter that defines the rate of convergence between the categories (See Azar and Menassa 2012 for additional details). The LI of Green occupants was assumed to be proportional to how far away the actual building performance is from the target one. Overconsumption of energy is expected to increase the LI of the Green occupants leading to higher rates of convergence to the Green category. As for the influence of non-Green people, it was randomly generated using a uniform distribution between 0 and the LI of Green people. This particular function was chosen to generate, in a random manner, a lower influence for non-green people. Table 1 illustrate the data exchange and time of engagement of the DOE2 and Anylogic models through the CERTI RTI federation (Menassa et al. 2013).

Table 1: Description of Steps Involved in Data Exchange within Federates and across the Energy Consumption Federation (adapted from Menassa et al. 2013)

a – A previously developed BIM of the building under analysis is manually loaded.
b - BIM is manually exported as an ifcXML file.
c - ifcXML file is transformed into an IMP file and imported in the DOE-2 model using the method described in Kim and Anderson (2012). The process is initiated manually.
d - DOE2 model is run for 12 months to generate the 'target' energy consumption levels and stored to be used in the DOE2 Federate for comparison with actual levels. Optimal occupancy behavior is assumed in this case. This whole process is automatically performed after launching the HLA model.
Step 0 - Anylogic is manually launched, which triggers the automated data exchange between DOE-2 and Anylogic through CERTI RTI Energy Simulation Federation.
Step 1 - Anylogic sends the initial number of Green and non-Green occupants to the HLA's ABM federate.
Step 2 - The ABM Federate communicates with the energy modeling Federate to translate the number of Green and non-Green to DOE-2 input parameters (INP file), which reflects the current behavior of building occupants. For instance, an increase in number of Green people results in lower lighting use schedules in DOE-2.
Step 3 and 9 - The energy modeling Federate launches the DOE-2 at t=0 or t-1, which predicts energy use for the first month (or second month).
Step 4 and 10- The energy consumption level for a given month is sent back to the energy modeling Federate which compares it to the 'target' energy level for this month that was generated in Step d.
Step 5 - The energy modeling Federate communicates the actual and target energy values to the ABM Federate.
Step 6 - The ABM Federate imports the values from Step 5 and launches a new run of the Anylogic model.
Step 7 - The updated values of Green and non-Green people resulting from changes in behavior due to Feedback are sent back to the ABM Federate. Changes in behavior resulted from the exposure of building occupants to their latest energy use levels (January) and their comparison to the desired levels (from Step d).
Step 8 - ABM Federate communicates the new numbers of Green and non-Green occupants to the energy modeling Federate in an INP file to be used as inputs in DOE2.

The steps keep repeating until the DOE-2 predictions correspond to the desired energy consumption levels, and the feedback intervention is considered successful. This provides the facility manager with an estimate of how long it will take for a certain level of feedback to influence occupants to reduce energy use. Results and Benefits of the Proposed HLA Framework: The results of the simulation are presented in Figure 2 (Menassa et al. 2013), where the graph on the left shows the behavior of occupants obtained from Anylogic and the graph on the right illustrates the building energy performance obtained from DOE-2. The split in behavior led to an actual energy consumption level that is 14 percent higher than the desired one (Figure 3, graph on the right). This overconsumption of energy resulted in a conversion of people towards the Green category over time. Convergence occurred at

month 37 with all occupants becoming Green, resulting in maximum energy savings. The results confirmed the successful coupling of DOE-2 and Anylogic.
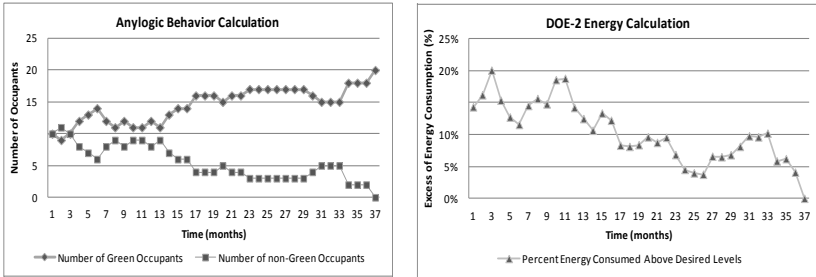
Figure 2: DOE-2 and Anylogic Results

CONCLUSION

The contributions of this research enable the coupling of building energy performance and operations for comprehensive and flexible energy simulation. The proposed framework will allow building stakeholders to use a simulation environment that models complexity as opposed to only approximating complexity. This is the case with most of other energy simulation frameworks discussed in this paper. Another important contribution lies in the HLA interface, which allows users to easily and effortlessly add additional software or hardware components depending on their intended applications and needs.

ACKNOWLEDGEMENTS

REFERENCES

Andrews, C. J. and Krogmann, U. (2009). "Explaining the adoption of energy efficient technologies in U.S. commercial buildings." Energy and Buildings, 41(3), 287-294.

ASHRAE. (2007). *55-2007, Thermal Environment Conditions for Human Occupancy*. American Society of Heating Refrigerating and Air-Conditioning Engineers Inc., Atlanta, GA.

ASHRAE (2010). "Energy Standard for Buildings except Low-Rise Residential Buildings." American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA.

Azar, E., and Menassa, C. C. (2012). "Agent-Based Modeling of Occupants' Impact on Energy Use in Commercial Buildings." *Journal of Computing in Civil Engineering*, ASCE, 26(4), 506.

Brooks, C., Lee, E. A., Liu, X., Neuendorffer, S., Zhao, Y. and Zheng,H. (2007). "Ptolemy II – heterogeneous concurrent modeling and design in Java." Technical report No. UCB/EECS-2007-7, Berkeley, CA: University of California at Berkeley.

Dahmann, J. S., R. M. Fujimoto, and R. M. Weatherly (1998), "The DoD High Level Architecture: An Update', Proceedings of the 1998 Winter Simulation Conference, IEEE.

EIA (2012). "U.S. Energy Information Administration Glossary." Energy Information Administration. Available at: http://www.eia.gov/tools/glossary/index.cfm, last accessed September 2012.

Hensen, J. (1999). "A comparison of coupled and decoupled solutions for temperature and air flow in a building." ASHRAE Transactions, 105 (2), 962–969.

Kim, H., and Anderson, K. (2012). "Energy Modeling System using Building Information Modeling (BIM) Open Standards". Journal of Computing in Civil Engineering, in press. Available at: doi: 10.1061/(ASCE)CP.1943-5487.0000215.

Kuhl, F., Weatherly, R., and Dahmann, J. (1999), "Creating Computer Simulation Systems: An Introduction to the High Level Architecture", Prentice Hall.

Lam, K.P., Mahdavi, A., Gupta, S., Wong, N. H., Brahme, R. and Kang, Z. (2002). "Integrated and distributed computational support for building performance evaluation." Advances in Engineering Software, 33 (4), 199–206.

Menassa, C.C., Kamat, V. R., Lee, S., Azar, E., Feng, C., and Anderson, K. (2013) "A Conceptual Framework to Optimize Building Operations by Coupling Energy Simulation Models with Energy Management Systems." Journal of Computing in Civil Engineering, ASCE (Submitted).

Pang, X., Wetter, M., Bhattacharya, P. and Haves, P. (2012). "A framework for simulation-based real-time whole building performance assessment." Building and Environment, 54 (2012) 100-108.

Peschiera, G. and Taylor, J. (2012). "The Impact of Peer Network Position on Electricity Consumption in Building Occupant Networks Utilizing Energy Feedback Systems," Energy and Buildings, 49 (June 2012), 584-590.

Peschiera, G., Taylor, J., and Siegel, J. (2010). "Response-Relapse Patterns of Building Occupant Electricity Consumption Following Exposure to Personal, Contextualized and Occupant Peer Network Utilization Data," Energy and Buildings, 42(8): 1329-1336.

Trecka, M., Hensen, J. and Wijsman, A. (2006). "Distributed building performance simulation – a novel approach to overcome legacy code limitations." International Journal of HVAC&R Research, 12 (3a), 621–640.

Wetter, M. (2011). "Co-simulation of building energy and control systems with the Building Controls Virtual Test Bed." Journal of Building Performance Simulation, 4(3), 185-203.

XJ Technologies. (2009). "Anylogic Overview." Available at: http://www.xjtek.com/anylogic/overview/, Last accessed October 2012.

Zhai, Z.J. and Chen, Q.Y. (2005). "Performance of coupled building energy and CFD simulations." Energy and Buildings, 37 (4), 333–344.

Agent-based Modeling of Impact of Stakeholder Requirements on Sustainable Retrofit of Buildings

Kristina Stephan, S. M. ASCE[1] and Carol C. Menassa, PhD, A. M. ASCE[2]

[1]M.S. Student, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 2231 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706, PH (608) 262-3542, FAX (608) 262-5199, email: kastephan@wisc.edu

[2]M.A. Mortenson Company Assistant Professor of Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 2318 Engineering Hall, 1415 Engineering Dr., Madison, WI 53706, PH (608) 890-3276, FAX (608) 262-5199, email: menassa@wisc.edu

ABSTRACT

This paper uses an agent-based model to illustrate how social networks and agent interactions can help stakeholders of an existing building prioritize their values of cost awareness, energy saving and organizational performance, to ultimately come up with an optimal building retrofit decision. The model is tested using a small-world network to see how the network structure affects the overall retrofit priority in a medium size building case study. The agents interact based on the bounded confidence model, which may result in continuous changes in their three values over time. The results indicate that although sustainable retrofits are generally decided on from economical and environmental perspectives, incorporating social values may help the owner choose an approach that can reduce energy consumption while maintaining tenant comfort and well being. Further, the time to reach convergence in values was highly dependent on the owner's initial values and how extreme these were relative to the other stakeholders.

INTRODUCTION

A truly sustainable retrofit decision in a commercial office buildings needs to account for environmental, economic and social aspects and how much each building stakeholder values these (Savitz and Weber 2006). This presents a great opportunity to investigate ways in which costs, energy use and emissions can be reduced through sustainable retrofits, while accounting for the comfort, health and productivity levels of building end-users (Heerwagen 2000). Although sustainable retrofits have been researched in the past, few studies have modeled how different types of building stakeholders with differing values can agree on the multiple aspects of a sustainable retrofit during the decision making process.

Several studies have investigated how stakeholders perceive sustainable retrofit requirements of a building. For example, Baer (2012) looked at how stakeholder perceptions of 30 social, economic, environmental and technical requirements affected a retrofit decision in a commercial building; however, stakeholder interactions were not considered. Others evaluated interactions amongst decision makers; however, not related to achieving optimal sustainable retrofits. Anumba et al. (2002) incorporated interactions

between stakeholders based on collaboration on construction projects. The paper showed how stakeholder interactions can help optimize the design and construction processes however costs and energy conservation were not considered.

A building social network connects stakeholders who may interact to convince one another to change how they value the retrofit process. This can eventually lead to a possible agreement on how a retrofit is valued to help achieve a comprehensive decision.

OBJECTIVES

The aim of this research is to develop a framework to study how stakeholders in a commercial building network can be influenced by other stakeholders they interact with during a retrofit decision process to change how much they value the cost of a retrofit, energy saving and occupant comfort. These were chosen because the cost of a project is always an integral part of assessment in the construction industry including construction costs as well as long-term operation and maintenance costs (Ding 2008). Energy saving is also key to minimize environmental impact and save resources (Ding 2008). Finally, organizational performance has been shown to be directly linked to user comfort and well being within the workspace (Heerwagen 2000). Although it is common on construction projects for the owner to be the main decision maker (Martinez-Moyano et al. 2011), stakeholders will place different values on each of these three aspects. Thus, the objectives of this research are to investigate: (1) How does the time to reach agreement between all stakeholders differ with varying the owner's initial values of cost, energy and comfort?; and (2) Can the owner's values of cost, energy and comfort change upon interacting with other stakeholders, even if he/she has extreme initial values?

METHODOLOGY

A three-step methodology is adopted to achieve the objectives of this research:
(1) Define Stakeholders and Networks: The stakeholders considered in this research are the owner, building manager, contractor or architect/engineer (A/E) and tenants. Mason et al. (2007) define three main network types (random, small-world and scale-free) that best represent social dynamic interaction and influence. Their main properties and examples in relation to the type of commercial buildings are given below:

1) *Random networks* assume all stakeholders are connected randomly with a given number of connections such as in a small office building where all tenants are from different departments of the same company.

2) *Small-world networks* connect all stakeholders with connections being either local or long-distance (Watts 1999), such as in a medium size building. These represent the closeness of relationship of those connected. For example, a building manager who responds to tenant requests may be closely involved with them as well as the owner who he needs to relay this information to (local). However, the contractor or A/E may interact with the building manager (local) to discuss design/construction changes while rarely interacting with tenants (long-distance).

3) *Scale-free networks* have some stakeholders with many connections to one main node, described as "hubs" while others have few connections, known as

"hermits" (Barabasi and Albert 1999). This represents a more flexible structure such as a large building, where for example, the "hubs" are the owner and the building manger that connect to the different companies, "hermits".

(2) Define Stakeholder Interactions in Networks: After setting up the network structure, it is important to illustrate the interaction between the different stakeholders to identify how decisions are made. During these interactions, the stakeholders will exchange information on how much they value the retrofit and will try to influence each other to adjust these values to eventually reach a possible agreement about what retrofits should be implemented. Each stakeholder will also have a level of confidence corresponding to each value that reveals how certain he/she is of his/her value. Stakeholders who are less certain of their values may be more easily convinced by others to adjust their values and vice-versa. This uncertainty provides a threshold that specifies the range in which the stakeholder can change their value, also known as the bound of confidence (BC) (Hegselmann and Krause 2002; Weisbuch et al. 2003). Combining all three values with the corresponding uncertainties creates an opinion space for each stakeholder, representing *all* values of the retrofit process. Each value and uncertainty may be initialized between [0 1] with a value closer to one representing a stakeholder who feels very strongly toward prioritizing that attribute of the retrofit and is more confident. Extending this to three dimensions is illustrated by a cube opinion space equal to [0 1]³, shown in Figure 1 below (Lorenz 2008). The maximum of [1 1 1] represents a stakeholder who zero uncertainty and feels strongly toward all three values.
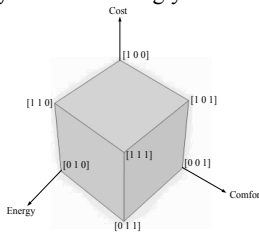
Fig. 1-Opinion space for all three values, adapted from Lorenz (2008)

Once the values and uncertainties are initialized, the interactions need to be modeled. These can be best evaluated using the relative agreement (RA) model, an extension of the BC model presented by Deffuant et al. (2002). The main difference between the BC and RA model is the way the RA model takes into account the uncertainty of each value. The change in value for a *stakeholder j* under the influence of a *stakeholder i* is proportional to the overlap between the agreement, divided by the uncertainty of the influencing segment, hence the term "relative". It is assumed that one cannot be influenced out of their range of BC (i.e., values are always between 0 and 1). Taking two stakeholders connected in the network, such as *stakeholder i* and *stakeholder j,* will each have three values; x representing cost, y representing energy and z representing comfort. Each of these values will also have corresponding uncertainties; Ux_i, Uy_i and Uz_i making up the threshold. Taking cost as an example, the value of *stakeholder i* will be compared to that of *stakeholder j* to find the overlap, O_{ij}. This overlap is graphically shown below in Figure 2 where, $O_{ij} = \min(x_i + Ux_i, x_j + Ux_j) - \max(x_i - Ux_i, x_j - Ux_j)$ [Eq. 1]. If

the overlap is found to be greater than the uncertainty, influence occurs. This means a change in value occurs for *tenant j* such that $x_j = x_j + \lambda \left[\left(\frac{O_{ij}}{Ux_i}\right) - 1\right] * (x_i - x_j)$ [Eq. 2] as well as uncertainty, $Ux_j = Ux_j + \lambda \left[\left(\frac{O_{ij}}{Ux_i}\right) - 1\right] * (Ux_i - Ux_j)$ [Eq.3], where λ is a constant parameter to control the intensity of the interactions (i.e., a higher λ may make the model converge faster vice versa).

Case where Oij > Uij
Influence occurs

Case where Oij < Uij
Influence does not occur
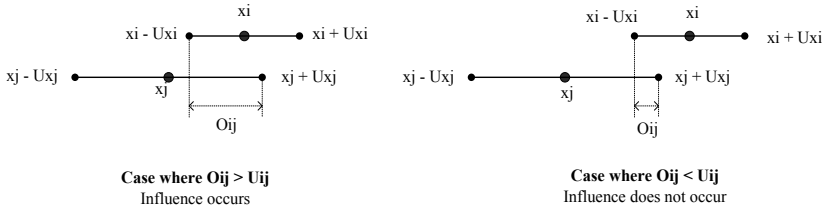
Fig. 2-Overlap between two stakeholder's values and uncertainties, adapted from Deffuant et al. (2002)

It is assumed that in order to agree with others on a retrofit decision upon interaction, a stakeholder may have to sacrifice his value of say cost, to account more for energy. This is known as the compensation effect and is because the model assumes the stakeholder judges distances in the 1-norm by looking at the sum of the absolute values in all dimensions (Lorenz 2007). A stakeholder is willing to accept a higher absolute value in one dimension if this is compensated by a lower absolute value of proportional magnitude in the other dimensions such that an increase, Δ, in one value will cause a decrease in the other two values by an equal amount of $\frac{\Delta}{2}$, or vice-versa.

The final step looks at how important a stakeholder is in influencing others, defined using Olander and Landin (2005) power/interest matrix which indicates that stakeholders who possess power to influence decisions in construction projects are liable to change depending on their level of interest in the decisions. Thus, for the purpose of this research, four types of stakeholders are identified as shown in Figure 3 below.
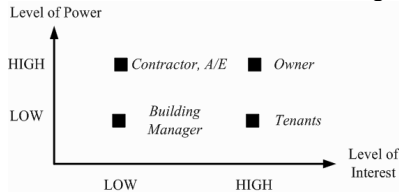
Fig. 3-Stakeholders in power/interest matrix

Applying this power/interest matrix was considered as a factor that updates the three values of *stakeholder j* after influence and compensation, such that $x_j = \sum_1^n factor_i * x_j$, $y_j = \sum_1^n factor_i * y_j$, $z_j = \sum_1^n factor_i * z_j$. According to the power/interest matrix in Figure 3, it is assumed in this research that the factors are as follows: owner = 1, contractor or A/E = 0.75, tenants = 0.5 and building managers = 0.25. After several iterations of interactions amongst all stakeholders, a point in time will be reached where

all stakeholders fall into each other's bounds of confidence and no influence can occur any further. All stakeholders would have converged to one retrofit decision.

(3) Modeling Tool and Model Description: Agent-based modeling (ABM) was selected to model the stakeholder interactions. It offers a way to model social systems that are composed of agents who interact with and influence each other, learn from their experiences and adapt their behaviors to better suit their environment (Macal and North 2010). AnyLogic 6.7.0 University was chosen to implement the model.

The agents represent the various stakeholders that may exist in a commercial office building (listed above). In this case, each tenant represents one company within the commercial building. They each have six variables representing the three values and the corresponding uncertainties. The interactions are then defined throughout the simulation depending on the social network chosen to represent the building. Over one time-step, all connected agents will interact through two-way communication. That is *agent i* will interact with *agent j* and then *agent j* will also interact with *agent i*.

The first step is for an *agent i* to contact all other agents in the network he/she is connected to. This triggers a function that calculates compares the three overlaps, O_{ij} for cost, energy and comfort, between both agents according to Eq. 1. The second step checks if the overlap for each value was found to be greater than the corresponding uncertainty of *agent i*. If this condition is found to be true, a sub-step updates that value and uncertainty for *agent j* according to Eq. 2 and 3 respectively. This update of each value then accounts for the amount of compensation that has occurred for the other two values. The third step multiplies these new values according to the factors in the power/interest matrix. This new value and uncertainty will now be used in the next interaction. The process is repeated until the convergence is reached, at which point the simulation stops. Convergence is continuously checked after a minimum of six time steps, allowing each *agent i* to contact and interact with all other agents he is connected to. If the variance of values for cost, energy and comfort for all stakeholders is less than the assumed value of 0.001, convergence is reached. This means the stakeholders no longer have influence on changing each other's values. The model then outputs the convergence time (i.e., number of time steps) of the overall network and individual times of each stakeholder (since some may converge before others).

CASE STUDY EXAMPLE

A case study of a medium size building is used to verify the results of the model. The Commercial Building Energy Consumption Survey (CBECS) (US Energy Information Administration 2003) was used to gather real data on medium sized office buildings in the US to obtain a realistic representation of how social networks may exist in these buildings which showed a building of medium size has three tenants (i.e., companies) on average. A small-world network was used as this best represents this size of building as discussed above. The network was characterized by N-1 connections, such that each person is connected to an average of N-1 other agents and is decentralized such that there is no dominant node to which other nodes directly connect (Watts 1999). A very high clustering coefficient, C, equal to 0.95 was assumed to show that most those within the network are strongly overlapping (i.e., we expect that 95% of the stakeholders

to interact at each time step). Three test cases were implemented. The first initializes the owner's three values to a uniform positive distribution, which is the base case. This means the owner randomly values cost, energy and comfort and has not particular preference. The second case initializes the owner's values to 0.9, an extreme owner with strong commitment to all values of sustainability. The final case takes the owner's initial values at 0.5, representing a neutral owner who is indifferent and may be more flexible to vary of his. Under all cases, the remaining stakeholders' values are initialized to a uniform positive distribution.

Analysis and Results: Statistical analysis is applied to the values obtained from running the model 1000 times for three cases of the owner's initial values described above.

Convergence times: After producing several histogram plots, it was found that the data did not follow normal distributions and therefore it was decided that the Kruskal-Wallis (KW) test was the most appropriate (Ruxton and Beauchamp 2008). The first step was to test the null hypothesis that the means of the maximum convergence time were the same under the three cases such that H_o: $\mu_{Uniform} = \mu_{0.9} = \mu_{0.5}$ and H_1: means are not equal. If the means were found to be unequal, the Wilcoxon test was then used to test whether the owner's original values are statistically significant. Under the KW test, the p-values were found to be extremely small which shows strong indication against H_0. Thus, statistical significance in the overall network convergence times under the three cases is observed. The Wilcoxon test was then used, applying the Bonferroni correction to the significance level to account for Type 1 error (Wilcox 2011) and ensure the overall error level over all three independent samples remains at the 5% level.

   Comparing the overall convergence times of the uniform positive initial values case to 0.9 and 0.5 was statistically significant with p-values of 2.2e-16 and 1.679e-14 respectively, much less than the 5% confidence level. When comparing the 0.9 case to 0.5, the change in convergence time was also statistically significant. This implies that changing the owner's initial value of cost, energy and comfort can greatly affect the results for how long the network overall takes to converge. Looking at the means for the overall convergence times, it was found that the uniform positive case, 0.9 and 0.5 took 440, 706 and 262 time steps respectively for all values of all stakeholders to converge. This correctly shows that if an owner begins with initial values that are extreme (i.e., 0.9), it will take longer for all stakeholders to reach convergence on retrofit options.

Difference between owner's original and final values at convergence: To track if the change in owner's values is statistically significant, the Wilcoxon test is also used, however the direction of change is specified, such that the H_0: D = 0 and H_1: D > 0 at a 5% confidence level. This hypothesis tests the difference, D = Value $_{x, y\ and\ z}$ (t = 0) − Value $_{x, y\ and\ z}$ (t = overall convergence time) to check if it is different than zero, to show if there is a statistically significant change from the original value to final. Three samples are tested independently for the owner's cost, energy and comfort values under the uniform positive, 0.9 and 0.5 cases. A significant change was only found when there was a positive difference. The p-values for the three values were all found to be extremely small and hence statistically significant under all three cases. This implies that the way the owner values cost, energy or comfort may change significantly after interacting with

other stakeholders regardless of his original values. The 0.5 case showed the highest change with an average drop of 0.4 drop in all three values, showing that a neutral owner is capable of becoming very flexible in his values.

Figure 4 below shows an example of one simulation run for the case of the owner having an original uniform cost value. Each of the three graphs shows all stakeholders' change in values over the first few time steps. These graphs help depict how compensation occurs from one-time step to another, such that if one stakeholder's value for cost increases, such as for Tenant 3, his/her value for energy and comfort will decrease. It can be concluded that even though the owner starts as a stakeholder with a neutral opinion (i.e., around 0.5 in this case), this is subject to change and convergence may still eventually be reached after several interactions, even though other stakeholders may become more extreme such as the case of Tenant 3.
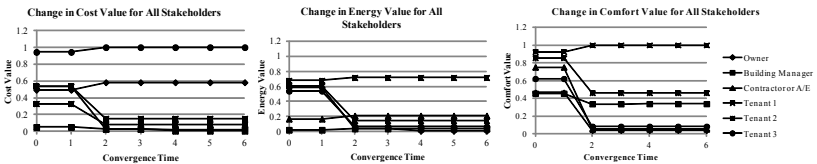
Fig. 4.-Graphs showing changes in all stakeholders' three values for one run

CONCLUSIONS

The ABM used in this research established a framework to allow incorporation of many stakeholders in the decision making process of sustainably retrofitting an existing building. The model shows interactions amongst stakeholders in a building social network can affect retrofit values of cost, energy and comfort. The uncertainty levels of each stakeholder's values are also critical in determining their willingness to alter their values and adjust toward or away from an influencing stakeholder's values. Evaluating the different cases showed that when the owner had an extreme initial value of 0.9, the convergence time was the longest, yet convergence was still reached. It was also shown that a neutral owner with an initial value of 0.5 was also able to agree with the stakeholders on retrofit value to eventually reach convergence. Future research is focused on studying how different social networks and building sizes affect the results.

AKNOWLDGEMENT

REFERENCES

Anumba, C. J., Ugwu, O. O., Newnham, L., and Thorpe, A. (2002). "Collaborative design of structures using intelligent agents." *Automation in Construction*, 11(1), 89–103.

Baer, B. (2012). "Sustainable Retrofits in Existing Buildings Analysis of Stakeholder Types, Requirements, and Perceptions in Decision Making, using the House of Quality." University of Wisconsin-Madison.

Barabasi, A.-L., and Albert, R. (1999). "Emergence of Scaling in Random Networks." *Science*, 286(5439), 509–512.

Deffuant, G., Amblard, F., Weisbuch, G., and Faure, T. (2002). "How can extremism prevail? A study based on the relative agreement interaction model." *Journal of Artificial Societies and Social Simulation*, 5(4).

Ding, G. K. C. (2008). "Sustainable construction - The role of environmental assessment tools." *Journal of Environmental Management*, 86(3), 451–64.

Heerwagen, J. (2000). "Green buildings, organizational success and occupant productivity." *Building Research and Information*, 28(5/6), 353–367.

Hegselmann, R., and Krause, U. (2002). "Opinion Dynamics and Bounded Confidence Models, Analysis, and Simulation." *Journal of Artificial Societies and Social Simulation*, 5(3).

Lorenz, J. (2007). "Repeated Averaging and Bounded Confidence Modeling, Analysis and Simulation of Continuous Opinion Dynamics." University of Bremen, Germany.

Lorenz, J. (2008). "Fostering Consensus in Multidimensional Continuous Opinion Dynamics under Bounded Confidence." *Managing Complexity: Insights, Concepts, Applications*, 321–334.

Macal, C., and North, M. (2010). "Tutorial on agent-based modelling and simulation." *Journal of Simulation*, 4(3), 151–162.

Martinez-Moyano, I. J., Zhao, F., Simunich, K. L., Graziano, D. J., and Conzelmann, G. (2011). "Modeling the Commercial Buildings Sector: an Agent-based Approach." *ASHRAE 2011 Annual Conference*, Montreal, QC, 366–374.

Mason, W. A., Conrey, F. R., and Smith, E. R. (2007). "Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence Within Social Networks." *Personality and Social Psychology Review*, 11(3), 279–300.

Olander, S., and Landin, A. (2005). "Evaluation of stakeholder influence in the implementation of construction projects." *International Journal of Project Management*, 23(4), 321–328.

Ruxton, G. D., and Beauchamp, G. (2008). "Some suggestions about appropriate use of the Kruskal–Wallis test." *Animal Behavior*, 76(3), 1083–1087.

Savitz, A. W., and Weber, K. (2006). *The Triple Bottom Line: How Today's Best-Run Companies Are Achieving Economic, Social, and Environmental Success and How You Can Too*. Jossey-Bass, 300.

US Energy Information Administration. (2003). *Commercial Buildings Energy Consumption Survey*.

Watts, D. J. (1999). "Networks, Dynamics and the Small-World Phenomenon." *American Journal of Sociology*, 105(2), 493–527.

Weisbuch, G., Deffuant, G., Amblard, F., and Nadal, J. P. (2003). "Interacting Agents and Continuous Opinions Dynamics." *Heterogenous Agents, Interactions and Economic Performance*, 521, 225–242.

Wilcox, R. R. (2011). *Introduction to Robust Estimation & Hypothesis Testing*. Elsevier, 690.

A multi-scale tunnel product model providing coherent geometry and semantics

A. Borrmann and J. Jubierre

Chair of Computational Modeling and Simulation, Technische Universität München, 80290 München; PH +49 89 289 23047; FAX +49 89 289 25051; email: {Borrmann, Jubierre}@bv.tum.de

ABSTRACT

For the planning of large infrastructure projects, such as inner-city subways, widely differing scales have to be considered – ranging from the scale of several kilometers for the general routing of the subway down to centimeter scale for the detailed planning of subway stations and hubs. At the same time the strongly inter-disciplinary character of these projects necessitates an intensive exchange of data between the individual stakeholders. Data exchange on a high semantic level is supported by the technology of product models. However, currently available product models do not support the multi-scale aspects required to properly handle large infrastructure models. This paper contributes to closing this technological gap by proposing a general methodology for developing multi-scale product models which combine semantic and geometrical aspects in a consistent and coherent manner. The methodology is illustrated by the development of a multi-scale product model for shield tunnels. In addition, methods for the automated preservation of the consistency of geometric representations at different scales are discussed.

INTRODUCTION

For the planning of large infrastructure projects, such as inner-city subways, widely differing scales have to be considered – ranging from the kilometer scale for the general routing of the carriageway down to the centimeter scale for the detailed planning of subway stations and hubs. Despite the multi-scale characteristics inherent in the planning of carriageways, today's infrastructure planning software does not support multi-scale geometric modeling. The research unit "3DTracks", funded by the German Research Foundation, is tackling this issue by developing a methodological basis for introducing multi-scale geometry into civil engineering models.

The concept of multiple geometric representations at different scales is well known from both cartographic applications as well as 3D city modeling. For example CityGML (Kolbe 2008), an open standard for the storage of 3D city models based on the Geographic Markup Language (GML), provides five different Levels of Detail (LoD). The LoD concept in these application areas relies on the independent storage of individual geometric models at each level of detail. As the dependency between the individual levels is not explicitly represented, inconsistencies can easily arise. This is

less problematic for geographic applications where data sets are rather static and only rarely subject to modifications. However, for the planning of large tunneling projects, a more robust approach is needed.

In (Bormann et al. 2012a) and (Borrmann et al. 2012b), the authors introduced the concept of multi-scale geometric modeling for tunnel design, which provides mechanisms for the automated preservation of consistency between the different LoDs. The concept relies on the explicit definition of dependencies between geometric elements and has been implemented using technologies provided by parametric CAD systems.

However, for comprehensive use throughout the entire design and engineering process, it is necessary to incorporate semantics into the model. Models that comprise both a semantic and geometric description are usually referred to as product models. Examples are the Industry Foundation Classes (IFC) for building design and the CIS/2 model for structural steel projects. These models are based on object-oriented principles and provide typing, inheritance, attributes and relationships, resulting in powerful mechanisms for describing semantics. Consequently, product models form a sound foundation for ensuring interoperability between different software products and between different stages of the construction project.

Based on preliminary work by (Yakubi et al. 2007) we are introducing a comprehensive product model for shield tunnels which fulfills the demands of the design and engineering of large infrastructure projects. The main emphasis of this paper is placed on integrating the semantic description with the multi-scale geometry approach discussed above. In the presented concept, the multi-scale approach also forms part of the semantic model, i.e. specific entities are only available at a particular LoD. The major challenge is then to achieve and maintain semantic-geometric coherence in the model (Stadler & Kolbe 2007; Clementini 2010), which means that geometric elements at a certain LoD are assigned to correct semantic elements on the same level.

POINT OF DEPARTURE: A SHIELD TUNNEL PRODUCT MODEL

The point of departure for our investigations is the development of a product model for shield tunnels. Although a first draft for a shield tunnel product model was provided in (Yakubi et al. 2007), it had to be adapted to the specific needs of our research, in particular with respect to the multi-scale modeling approach. Figure 1 shows an overview of the developed model. The semantic model presented is aligned with the Industry Foundation Classes[1] (IFC), a comprehensive, standardized product model for buildings. In particular, we make extensive use of the space structure concept. As explained in detail below, it significantly simplifies the integration of multi-scale concepts into the model. Like the IFC model, the proposed tunnel product model provides a clear separation between semantic objects and the associated geometry.

---

[1] Please note that, for improved readability, we omit the prefix IFC in the class names throughout this paper.

The left-hand side of Figure 1 shows the semantic part of the model. Please note that, like in the IFC, we distinguish space objects (depicted in blue) from physical objects (depicted in green). The meaning of the individual entities is illustrated in Figure 2. Except for the ring space, all space objects represent longitudinal spaces along the entire *TunnelPart*. The *Ring* space, however, has the length of a segment only. The relations between the semantic objects rely on the space structure concept, modeling aggregation relationships between the site, the tunnel, the tunnel parts, the longitudinal spaces, and the rings.

The associated geometry representations are depicted on the right-hand side of Figure 1. The tunnel object is associated with a dedicated *Alignment* object. Since the alignment plays a key role in the design and engineering of tunnels, it is essential to provide the genuine alignment objects such as lines, arc segments and clothoids as part of the product model. Due the limited space available here, we have not depicted all these elements in detail here. For describing the geometry of the longitudinal objects (both the space objects and the physical objects), we rely on a profile geometry representation[2] to define volumetric geometry by means of cross-sections extruded along a given axis.
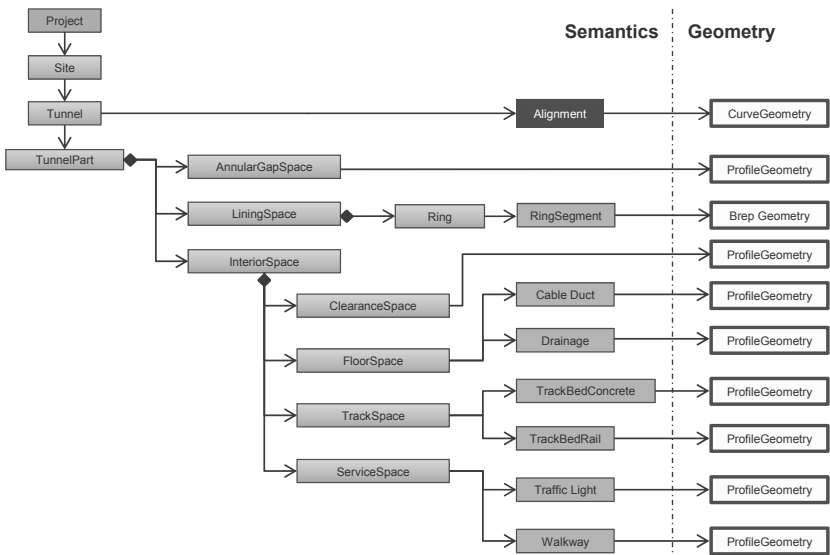
Figure 1: The proposed shield tunnel product model using conventional, non-multi-scale approaches[3]

---

[2] The IFC provide the entity IfcSurfaceCurveSweptAreaSolid to describe geometry by means of a closed profile and an extrusion along a reference curve.

[3] Please note that we use a simplified notation here. Relationships which are represented by dedicated objects in IFC are depicted here as direct associations. Aggregation relationships (IfcRelAgggregates) are marked with a rhombus. The containment relationships between Space Structure entities and physical objects (IfcRelContainedInSpatialStructure) are depicted as arrow lines.
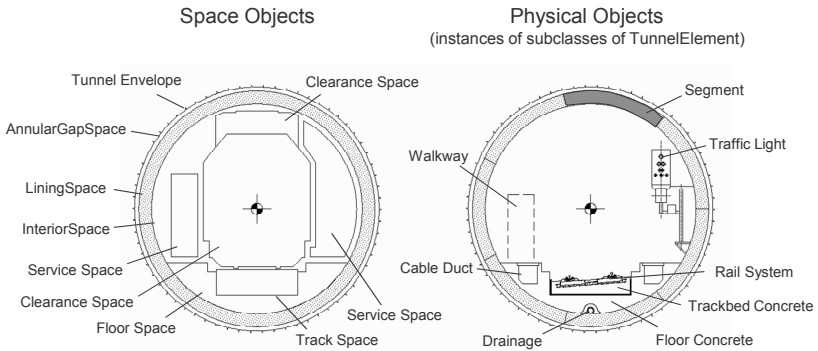
Figure 2: A tunnel cross-section showing the individual spaces (left) and elements (right) of the product model

INTEGRATING MULTIPLE SCALES INTO THE MODEL

A main shortcoming of existing product models is the lack of support for different levels of detail. As a consequence, they only poorly fulfill the demands of engineering practice, particularly in the early design phases, where 'coarse-grain' design information is successively enriched by more fine-grained information.

To better support the design and engineering of tunnels, we have developed concepts for integrating multi-scale approaches with tunnel product models. The main difference between our approach and the one followed by GIS standards, such as CityGML, is the scale-aware sub-division of the semantic part of the model. While the GIS standards allow the association of multiple geometric representations for the individual levels with one semantic object, but keep the semantic object structure fixed across the different LoDs, we propose to explicitly represent refinement relationships in the semantic part of the model, thus providing a much higher degree of semantic-geometrical coherence of the multi-scale model.

The resulting multi-scale product model is depicted in Figure 3. In order to group and provide access to all elements at a certain level of detail, we introduce dedicated LoD objects. These objects aggregate all spatial and physical objects at the corresponding level. At the same time, we maintain the aggregation relationships across the different LoDs in order to explicitly model a refinement hierarchy. One of the key aspects of our approach is that the refinement hierarchy is created with the help of space objects, while physical objects form part of the finest level only. This allows us to use spaces as placeholders on coarser levels, thus providing full compliance with standard IFC modeling approaches for space-element aggregation structures.

The geometry representation is basically identical to that of the model defined above. Figure 2 provides a 2D graphical illustration of the representations at the different LoDs, while Figure 4 provides a 3D illustration. Please note that on LoD1 the tunnel is represented by its axis only. On LoD2 the additional space object

TunnelEnvelope has been introduced to provide a semantic object representing the entirety of the tunnel. The *Ring* space objects belong to the finest level of detail, LoD5, since their definition happens at a more advanced stage of the planning process. Each *Ring* contains the *RingSegments* which belong to it.
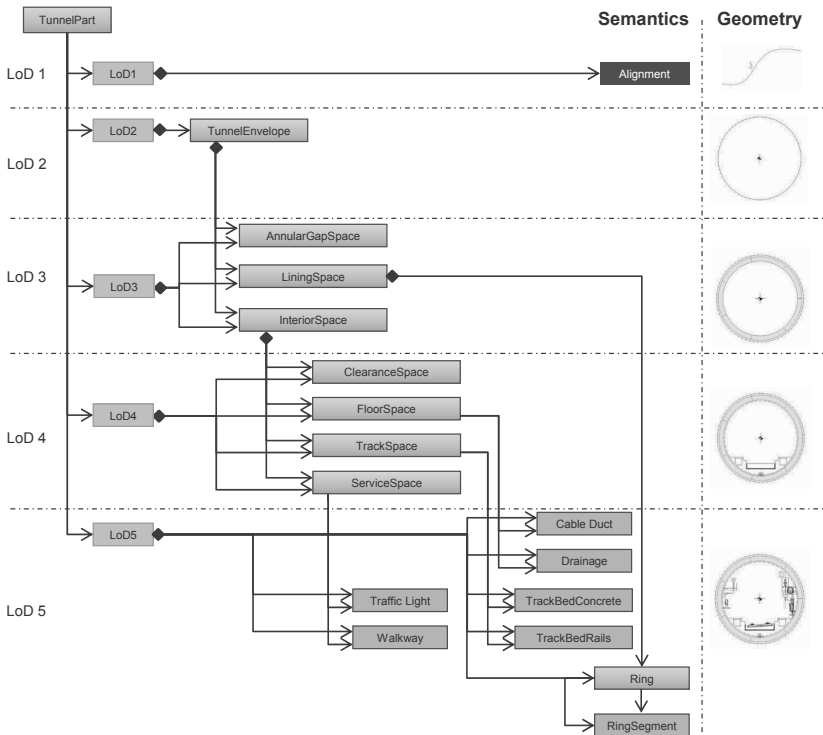
Figure 3: The proposed shield tunnel product model incorporating a coherent multi-scale representation of semantics and geometry

Although this model implements a multi-scale approach and provides a coherent representation of semantics and associated geometry, it does not yet provide a means of preserving the consistency between the different LoDs. This is caused by the fact that the geometry representations of the individual LoDs are independent of each other. Inconsistencies can arise, for example, when a modification is performed on one level, but not propagated to the other levels. To overcome this deficiency we propose to make use of procedural geometry descriptions, which allow us to explicitly define dependencies between individual geometric objects and thus provides a means for automatic consistency preservation. The resulting data model is described in the next section.
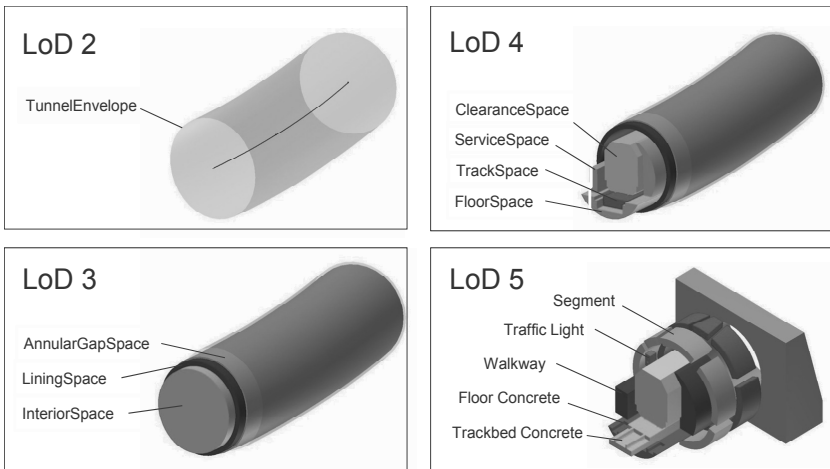
Figure 4: A 3D representation of the different LoDs of the multi-scale shield tunnel product model

PROCEDURAL GEOMETRY FOR AUTOMATED CONSISTENCY PRESERVATION

In the multi-scale product model introduced above, as well as in all multi-scale approaches known from the GIS domain, the geometry representations of the individual LoDs are stored independently of one another. In the case of modifications on one level, all other levels have to be updated manually in order to maintain the consistency of the entire multi-scale model. This is appropriate for geographic applications where we are faced with rather static models, which are rarely subject to modifications. However, for the highly dynamic processes of the design and engineering of tunnels, an automated means of preserving consistency is desirable.

This is realized by explicitly modeling dependencies between the geometric representations at the different LoDs using the concept of procedural geometry (Borrmann et al. 2012a, 2012b). The concept of procedural geometry is well known from parametric CAD systems, which usually provide parametric sketches allowing users to define constraints between individual geometric entities, and a construction history that relates individual construction operations to each other (Pratt 2010).

Figure 5 depicts how this form of geometry representation can be integrated into a multi-scale product model. The explicit geometry representation of individual elements of the model is replaced by a procedural one which can be linked to the geometric entities at lower LoDs. One example is the alignment curve which acts as the LoD1 representation and is subsequently used as the path for creating the extrusion geometry of all longitudinal objects on the finer LoDs. Another example is the sketch-based creation of profiles on finer levels from coarser ones using offset operations.

Applications that are capable of interpreting and processing procedural geometry are able to automatically preserve the consistency of the multi-scale model

by propagating changes on geometric objects to all dependent representations and updating them accordingly.

To prove the suitability of the concept, we have developed import and export modules for the parametric CAD systems Autodesk Inventor and Siemens NX, respectively. Using these modules, we are able to exchange inherently consistent multi-scale models between these CAD systems by means of the neutral product model presented in this paper.
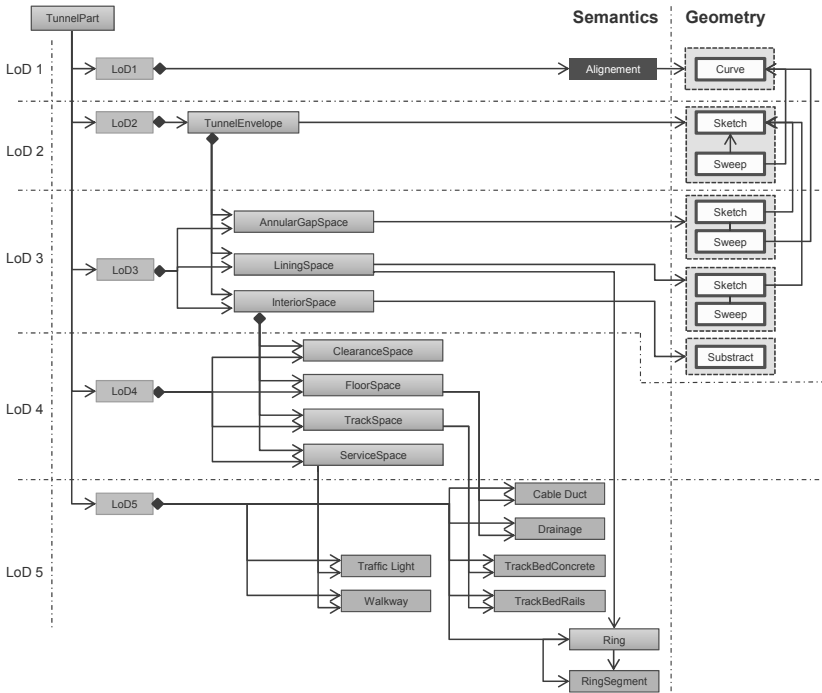
Figure 5: The shield tunnel product model incorporating a multi-scale semantic representation and a procedural geometry representation. It makes it possible to define dependencies between the geometric representations at the different LoDs, thus providing a means for preserving consistency across the levels.

CONCLUSION

This paper has introduced a novel approach for the design of multi-scale product models that provides inherent coherence between the semantic and the geometric entities. This approach was illustrated by describing the elaboration of a multi-scale product model for shield tunnels. The key aspect of the developed approach is the use of space aggregation hierarchies for representing refinement relationships on the semantic side. Physical objects which are located in these spaces are only included in

the finest level of detail. Each of the spatial and physical objects is associated with a corresponding geometry representation. The resulting product model represents a complete and spatio-semantically coherent description of a multi-scale product model for data exchange purposes. However, since the geometry representations are stored independently of one another, their consistency cannot be automatically preserved. To rectify this, the product model has been extended in a second step through the inclusion of procedural geometry descriptions which establish explicit dependencies between the geometry representations at the different levels of detail. The extended product model makes it possible to exchange multi-scale models that incorporate rules for the automated preservation of the consistency of the geometry representations of the individual levels of details. The feasibility of the concept has been proven by the development of import and export modules for the exchange of multi-scale models between two different modeling systems.

ACKNOWLEDGMENTS

REFERENCES

Borrmann, A., Ji, Y., Jubierre, J. (2012a): Multi-scale geometry in civil engineering models: Consistency preservation through procedural representations. In: *Proc. of the 14th Int. Conf. on Computing in Civil and Building Engineering*. Moscow, Russia

Borrmann, A., Ji, Y., Jubierre, J. (2012b): Procedural Modeling: A new approach to multi-scale design in infrastructure projects. In: *Proc. of the EG-ICE Workshop on Intelligent Computing in Civil Engineering*, Herrsching, Germany

Clementini, E. (2010): Ontological impedance in 3D semantic data modeling, 5th International 3D GeoInfo Conference, Berlin, Germany

Ji, Y., Borrmann, A., Beetz, J., Obergriesser, M. (2013): Exchange of Parametric Bridge Models using a Neutral Data Format, *ASCE Journal of Computing in Civil Engineering*, accepted, DOI: 10.1061/(ASCE)CP.1943-5487.0000286

Kolbe, T. H. (2008): Representing and Exchanging 3D City Models with CityGML. In: *Proceedings of the 3rd International Workshop on 3D Geo-Information*, Seoul, Korea

Pratt, M. J. (2010) Exchanging history-based parametric CAD models using ISO 10303. International Journal of Product Lifecycle Management, 4(4).

Stadler, A., Kolbe, T. H. (2007): Spatio-semantic coherence in the integration of 3D city models In: *Proceedings of the 5th International Symposium on Spatial Data Quality*, Enschede, The Netherlands

Yabuki, N., Azumaya, Y., Akiyama, M., Kawanai Y., Miya, T. (2007): Fundamental Study on Development of a Shield Tunnel Product Model. *Journal of Civil Engineering Information Application Technology* 16, pp. 261-268

Preliminary Model Development for Predicting Strength and Stiffness of Cement-Stabilized Soils Using Artificial Neural Networks

O. Wang[1] and A. Al-Tabbaa[2]

[1]PhD Candidate, Department of Engineering, University of Cambridge, Trumpington St., Cambridge, UK, CB2 1PZ; PH +44 (0) 1223 748586; email: ow226@cam.ac.uk
[2]Reader, Department of Engineering, University of Cambridge, Trumpington St., Cambridge, UK, CB2 1PZ; PH +44 (0) 1223 332715; email: aa22@cam.ac.uk

ABSTRACT

This paper presents ongoing work on data collection and collation from a large number of laboratory cement-stabilization projects worldwide. The aim is to employ artificial Neural Networks (ANN) to establish relationships between variables, which define the properties of cement-stabilized soils, and the two parameters determined by the Unconfined Compression Test, the Unconfined Compressive Strength (UCS), and stiffness, using E_{50} calculated from UCS results. Bayesian predictive neural network models are developed to predict the UCS values of cement-stabilized inorganic clays/silts as well as sands as a function of selected soil mix variables, such as grain size distribution, water content, cement content and curing time. A model which can predict the stiffness values of cement-stabilized clays/silts is also developed and compared to the UCS model. The UCS model results emulate known trends better and provide more accurate estimates than the results from the E_{50} stiffness model.

INTRODUCTION

Soil stabilization, a process which involves the introduction of cementitious agents to the native soil, either in the form of dry powder or slurry, has become a popular ground improvement and remediation method for improving soil engineering properties to overcome design and construction difficulties in poor quality ground. Cement, the most traditionally used agent, is often utilized in such circumstances, possibly with blends and additives. The design of cement-stabilized soil is most commonly based on results from the unconfined compression strength (UCS) test due to its simplicity and cost-effectiveness. The mechanical properties such as UCS and stiffness of cement-stabilized soils are functions of a number of variables, including the ones relating to the soil, binder, mixing and curing conditions, which collectively play a role in the performance of soil-cement systems. According to Terashi (1997), there are numerous factors that influence the properties of cement-treated soils, which are categorized into: (i) characteristics of soil, (ii) characteristics of binder, (iii)

mixing conditions, and (iv) curing conditions. Cement-mixed soil systems are usually assessed in the laboratory to investigate the effects of important variables on design parameters prior to their full-scale field application. Depending on the application, there are various design criteria to be considered including strength, stiffness/compressibility, strain at failure, tensile strength, leachability and permeability. However, strength and stiffness are the main design criteria for most applications.

The potential for using Artificial Neural Networks (ANN) as a predictive and design tool in geotechnical engineering applications has been demonstrated in a number of studies over the last decade or so (Goh, 2001; Shahin et al., 2001; Stegemann, 2001). More recently, ANN techniques have been applied to the development of predictive models for deep soil mixing (DSM) applications. Shrestha and Al-Tabbaa (2011) investigated the feasibility of using ANNs for DSM application modelling and demonstrated that ANNs are capable of providing reasonable UCS predictions. Shrestha (2012) developed ANN models for predicting UCS of laboratory cement-mixed clays and peats based on 519 cases of 39 different clays and 111 cases of 11 peats. Results from her study showed that reasonable predictions were made and were in good agreement with experimental measurements. Trani (2012) developed models for the prediction of both the UCS and the settlement of deep mixed soft ground and further confirmed the predictive capability of ANN models for DSM applications. This study aims to extend ANN modelling cement-soil stabilization applications and investigates the applicability of using ANNs for predicting stiffness of cement-stabilized clays and silts, as well as UCS of stabilized sands.

THE ANN STRUCTURE AND DATA DESCRIPTION

The ANN models in this study were developed using the computer-based software 'Neuromat's Model Manager' (Sourmail, 2004). The neural network utilized by this software is a feed-forward error back propagation network which underlies a Bayesian framework (Mackay, 2003). Predictions from Bayesian neural network models are associated with error bars which allow for certain limits to the uncertainty of the results. The models constructed in this work consist of a number of processing elements arranged in three layers: (a) an input layer, (b) a hidden layer and (c) an output layer. The input units are multiplied by adjustable connection weights and summed with a constant, and the units in the hidden layer take the information from the input layer and then feed it into the non-linear hyperbolic tangent function to produce the outputs. In order to minimize the errors associated with the computed output values, a training process is conducted during the model development procedure. The training is done through successive iterations of the training data by propagating the errors backward through the network to adjust the values of the weights between layers.

Two datasets from two categories of cement-stabilized soils were used in this study. The first dataset consisted of 219 data cases from a number of projects involving laboratory-prepared inorganic silty clays stabilized using Portland cement, either by dry or wet mixing (Bergado et al., 2005; Hassan, 2007; Zhu et al., 2007; Xiao, 2009; Jergandan, 2010), while the second dataset was composed of 223 data

cases of cement-mixed sands (Filz et al., 2005; Lewsley, 2008; Yoon and Abu-Farsakh, 2009; Ajorloo, 2010; Lapointe, 2012). The parameters employed as input units in the models are variables related to soil properties, cement characteristics, mixing and curing conditions. Both datasets were trained and validated for UCS models, and the first dataset was also used for stiffness (expressed as the secant modulus of elasticity E_{50}) modeling. The structure of the cement-stabilized silty clay and sand ANN models are illustrated in Figure 1(a) and (b). Table 1 summarizes the range of variables employed for the training and validation of the models.
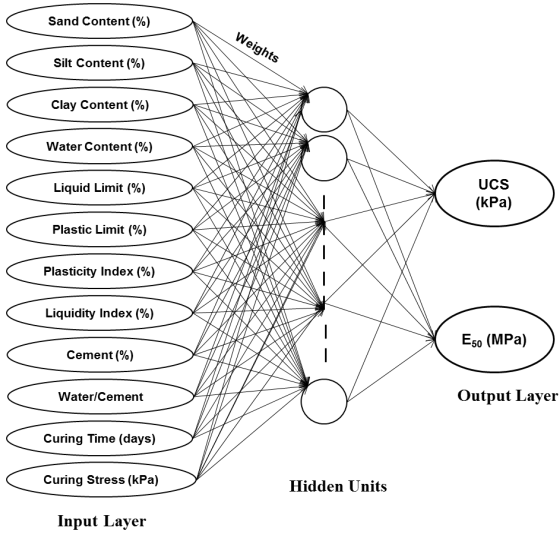
(a)
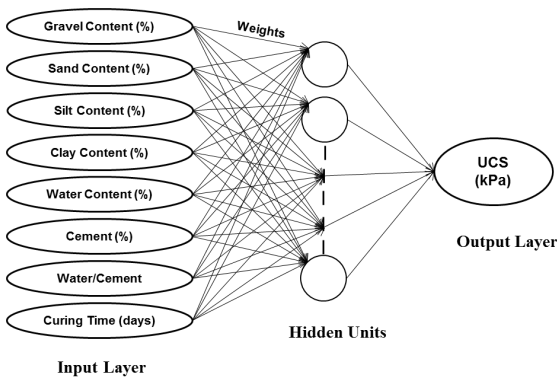
(b)

Figure 1. ANN model structures (a) cement-stabilized clay models; (b) cement-stabilized sand model

Table 1.Range of the values of model variables used in the study

Variables	Silty Clay Models				Sand Model			
	Mean	Min	Max	S.D.*	Mean	Min	Max	S.D.*
Gravel Content (%)	-	-	-	-	2.1	0	18	4.6
Sand Content (%)	7.3	0	43	8.7	70	20	100	26.7
Silt Content (%)	46.1	0	89	21.1	23.6	0	69.1	23.3
Clay Content (%)	46.5	10	100	22.4	4.4	0	33.2	7.2
Water Content (%)	99.1	27	250	41.1	23.2	7	72.5	13.4
Liquid Limit (%)	77.6	20.5	120	26	-	-	-	-
Plastic Limit (%)	34.1	14.7	57	8.5	-	-	-	-
Plasticity Index (%)	43.6	5.8	74.6	19.1	-	-	-	-
Liquidity Index (%)	1.6	0.6	3.1	0.6	-	-	-	-
Cement Content (%)	20	2.5	60	15.1	14.8	1	134.1	17.9
Water/Cement Ratio	7.8	1	32	5.6	2.7	0.3	15.8	2.1
Curing Time (days)	37	3	730	72	54	4	3044	286
Curing Stress (kPa)	12.6	0	500	55	-	-	-	-
UCS (kPa)	815	19	3706	690	1792	50	6760	1533
E_{50} (MPa)	101	1	1120	118	-	-	-	-

*S.D. – Standard Deviation

RESULTS AND DISCUSSIONS

Before training took place, each dataset was randomized and divided into three statistically consistent subsets of data: (i) training, (ii) testing and (iii) validation. According to Stegemann (2001) and Shahin et al. (2001), it is reasonable to have 80% of the data used for model training and testing, and 20% for model validation.

Cement-stabilized Silty Clay UCS and E_{50} ANN Models

Amongst the 219 data cases, 176 cases were used for calibration (88 individual cases each for training and testing) and 43 cases for validation. A total of 112 sub-models were trained and ranked based on their performance on predicting the output values on the unseen test data for both UCS and E_{50} ANN models. In order to obtain better results, a committee of sub-models was formed by selecting the combination of 'sub-models' that performed best (i.e. produced the minimum combined test error). In this case, the minimum combined test error was achieved when the committee consisted of 15 sub-models for the UCS model and 11 sub-models for the E_{50} model (Figure 2a and 2b). The committee was then retrained in order to obtain the best possible predictive ability.

Figure 3(a) and 4(a) depict how the predicted values of UCS and E_{50} match with the actual measurements, and Figure 3(b) and 4(b) show the comparison between the predicted and measured values for the unseen validation dataset. It can be

seen that the output values predicted by the UCS model matched closely with the measured values for calibration dataset, and the model was able to generate accurate predictions (validation dataset) based on high correlation coefficient (i.e. $r = 0.9322$). The coefficient of correlation between the measured and predicted values is a good indicator of the prediction performance of the model. Some dispersion from the 1:1 line can be noted in Figure 3(a) and (b). The reason behind this can perhaps be explained by the fact that the value of one or more input variables associated with these cases falls into the regions of very sparse data. The E_{50} model, on the other hand, predicted relatively less reliable values compared to the UCS model. Although a linear fit having a correlation coefficient value greater than 0.8 is generally considered to have strong correlation, dispersions and scatterings from the 1:1 line can be clearly observed in the E_{50} model, especially in the validation dataset where the model tended to overestimate the output values. The main reason behind the difference in performance between the two models can be perhaps attributed to the fact that the E_{50} values were determined by manually taking the slope of the line extending from the origin of the stress-strain plots to the point corresponding to half of the maximum stress, as these values were not available in the form of direct measurement from the UCT and none of them used on-sample measurements. If on-sample instrumentation was employed for the stiffness measurements, the E_{50} results would have been much more accurate, and that the E_{50} models could have produced more accurate and reliable predictions.

(a)   (b)
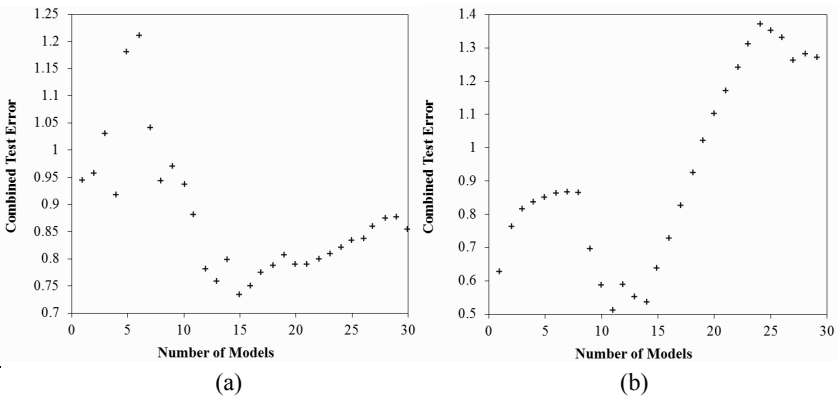
Figure 2. Test error vs. number of models in the committee (a) UCS; (b) E_{50}
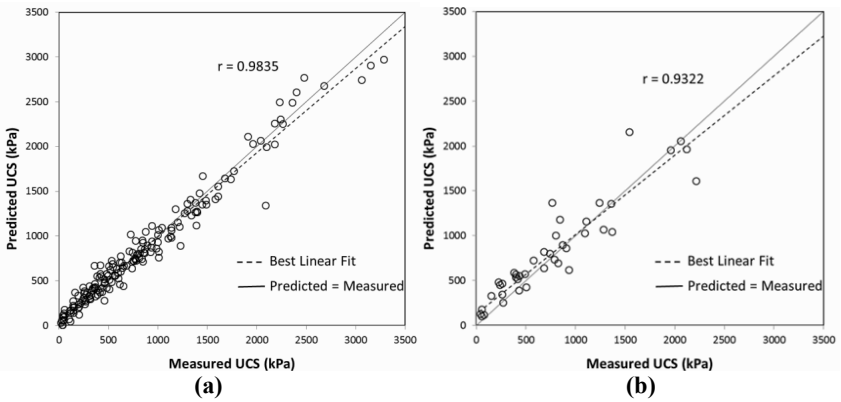
Figure 3. Predicted UCS values of cement-stabilized clays vs. actual measurements for (a) training and testing datasets; (b) unseen validation set
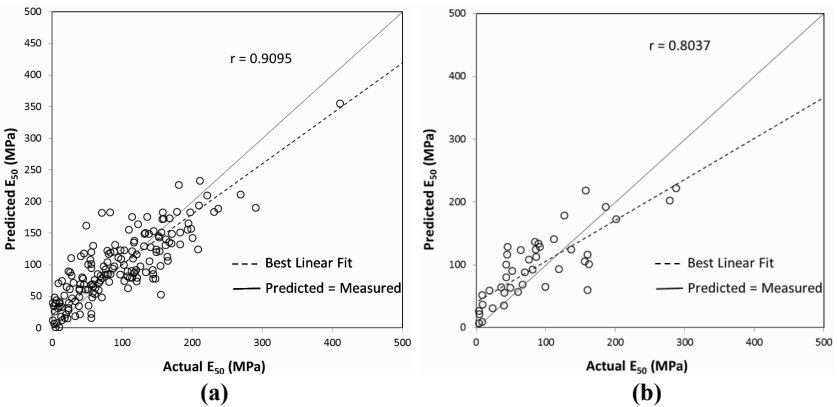
Figure 4. Predicted E_{50} values of cement-stabilized clays vs. actual values for (a) training and testing datasets; (b) unseen validation set

Cement-stabilized Sand UCS ANN Model

There were 180 out of 223 data cases used for training and testing, and 43 data cases for validation. A total of 123 sub-models were trained and tested for this model, and a committee consisting 15 sub-models was chosen and retrained. While Figure 4(a) shows the comparison between the predicted and measured UCS values for the training and testing datasets, the performance of the model on validation dataset is shown in Figure 4(b). The UCS values predicted by the model lied in close agreement with the actual measurements on the training and testing datasets, although some dispersion from the 1:1 line can be observed. When the model was assessed using the validation dataset, the majority of the predicted values matched closely with the measured UCS values in a reasonable manner, even though a much lower correlation

coefficient of 0.7699 was obtained compared to that from the calibration dataset. The dispersions observed from both plots could be possibly explained by the fact that these particular cases of cement-sand mix were the ones where a small amount of silica fume was added to the mix, while the other cases were only Portland cement mixed. It has been well recognized that silica fume is highly reactive and that the addition of silica fume can greatly improve strength of a cement-soil mix due to its pozzolanic properties and extreme fineness.
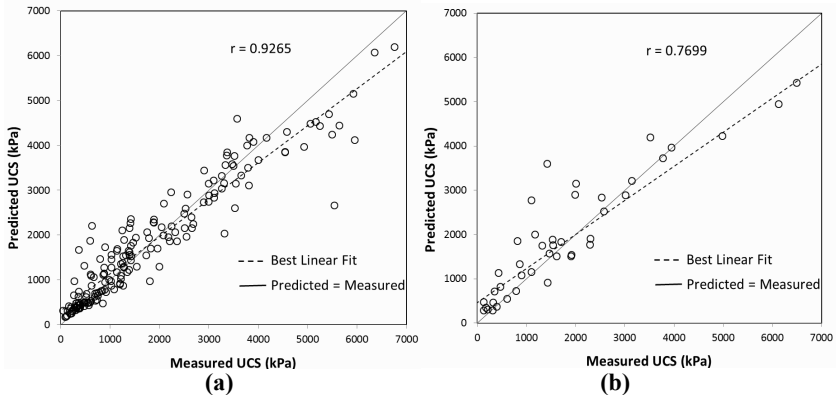
Figure 5. Predicted vs. measured values of cement-stabilized sand UCS model for (a) training and testing datasets; (b) validation set

CONCLUSION

This paper introduces current work in collecting and collating data from a large number of different laboratory projects on cement-stabilized soils, and developing predictive models using artificial neural networks (ANNs). Results in this study show that the reasonable UCS predictions made by the ANN models of cement-stabilized soils are in close agreements with experimental measurements. The cement-mixed clay UCS model indicates a higher predictive capability when compared to the stabilized sand model. Stiffness (E_{50}) ANN model of stabilized clay is also developed and compared to the UCS model, and it is concluded that the UCS model is able to provide more accurate predictions than the stiffness model, when the stiffness is determined using the stress-strain plot from UCS test data. On-going work of data collation and modeling is expanding the database and improving predictions as well as incorporating triaxial test data into the predictions and comparisons.

REFERENCES

Ajorloo, A.M. (2010). "Characterization of the mechanical behavior of improved loose sand for application in soil-cement deep mixing." *Ph.D thesis*. Université de Lille, Lille, France.
Bergado, D.T., Lorenzo, G.A., Taechakumthorn, C. and Balasubramaniam, A.S. (2005). "Compression behavior of high water content cement-admixed clay."

Proc. of Deep Mixing'05 - Intl. Conf. on Deep Mixing Best Practice and Recent Advances, 221-230.

Filz, G.M., Hodges, D.K., Weatherby, D.E. and Marr, W.A. (2005). "Standardized definitions and laboratory procedures for soil-cement specimens applicable to wet method of deep mixing." *GSP 136 Innovation in Grouting and Soil Improvement*, 1643-1656.

Goh, A.T.C. (2001). "Neural network applications in geotechnical engineering." *Scientia Iranica*, 8(1), 1-9.

Hassan, M. (2007). "Engineering characteristics of cement-stabilised soft Finnish clay- A laboratory study." *Master's Thesis*, Helsinki University of Technology, Finland.

Jegandan, S. (2010). "Ground Improvement with conventional and novel binders." *Ph.D thesis*, University of Cambridge, UK.

Lapointe, E. (2012). "Cement-treated soil: a comparison of laboratory and field data from fountain slide remediation deep mixing project." *MSc. thesis*, University of British Columbia, Vancouver, Canada.

Lewsley, G. (2008). "On the strength of saturated cement-treated soil reconstituted by wet-mixing." *MSc. thesis*, University of British Columbia, Vancouver, Canada.

Mackay, D.J.C. (2003). *Information theory, inference and learning algorithms*, Cambridge University Press, Cambridge, UK.

Shahin, M.A., Jaska, M.B., and Maier, H.R. (2001). "Artificial neural network applications in geotechnical engineering." *Australian Geomechanics*, 36(1), 49-62.

Shrestha, R. (2012). "Deep soil mixing and predictive neural network models for strength prediction." *PhD thesis*. University of Cambridge, Cambridge, UK.

Shrestha, R. and Al-Tabbaa, A. (2011). "Introduction to the development of an information management system for soil mix technology using artificial neural networks." *Geo-Frontiers 2011- Advances in Geotechnical Engineering Dallas, Texas: GSP* ASCE, Publication No 211, 576-584.

Sourmail, T. (2004). *Neuromat model manager manual*. Neuromat Ltd.

Stegemann, J.A. (2001). "Neural network analysis of the effects of contaminants on properties of cement pastes." *Ph.D. thesis*. Imperial College, London, UK.

Terashi, M. (1997). "Deep mixing method – Brief state-of-the-art." *Proc. of 14th Intl. Conf. on Soil Mechanics and Foundation Engineering*, 2475-2478.

Trani, L.D. (2012). "Application of artificial neural network to soft ground improvement: a case study." *Proc. of the Intl. Conf. on Ground Improvement and Ground Control 2012*, Wollongong, October, 465-470.

Xiao, H. (2009). "Yielding and failure of cement treated soil." *Ph.D thesis*, National University of Singapore, Singapore.

Yoon, S. and Abu-Farsakh, M. (2009). "Laboratory investigation on the strength characteristics of cement-sand as base material." *KSCE Journal of Civil Engineering*, 13(1), 15-22.

Zhu, W., Zhang, C.L. and Chiu, A.C.F. (2007). "Soil-water transfer mechanism for solidified dredged materials." *Journal of Geotechnical and Geoenvironmental Engineering*, 133(5), 588-598.

A Numerical Approach to Simulate Soil Freezing and Frost Heave behind an Earth Retaining Structure

Ming Zhu[1] Ph.D., P.E., M.ASCE and Radoslaw L. Michalowski[2], Ph.D., F.ASCE

[1]Senior Engineer, Geosyntec Consultants, 1255 Roberts Boulevard NW, Suite 200, Kennesaw, GA 30144; PH (678) 202-9500; email: mzhu@geosyntec.com
[2]Professor, Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109; PH (734) 763-2146; email: rlmich@umich.edu

ABSTRACT

Design and construction of earth retaining structures in cold regions requires special consideration for potential impact of frost heave. During the freezing process, additional earth pressure can be generated by frost heave as a result of the formation and growth of ice lenses in soil. The additional earth pressure due to frost heave may cause unacceptable deformation of the structures or even failure, if it is not properly considered in the design. This paper presents a numerical approach to simulate the freezing process and associated frost heave in soil. In this approach, soil is treated as a mixture of solid particles, unfrozen water, and ices. The expansion of soil due to frost heave is described mathematically using the Porosity Rate Function, where the porosity rate (i.e., change of soil porosity over time) depends on the water transport into the freezing zone, which is governed by the soil temperature, temperature gradient, and stress. The heat transfer and frost heave models have been incorporated into a finite element program for numerical simulations. The numerical approach was used to evaluate a concrete cantilever retaining wall under freezing conditions; the simulation yielded realistic results of ground and wall deformation.

INTRODUCTION

Cold regions encompass permafrost and seasonal frost zones and account for approximately half of the land surface of the earth (Freitag and McFadden, 1996). Low temperatures can have significant impact on soil properties and therefore, need to be considered in geotechnical engineering practice. One phenomenon that often occurs in cold regions is *frost heave*, which is indicated by the heaving of ground surface when it is subject to freezing. It has been discovered that frost heave is caused by the process called *ice lensing* (e.g., Taber, 1930 and Motou et al., 1998). When the ground temperature drops below the freezing point of water, the pore water in soil freezes and turns into ice crystals. The freezing process creates the cryogenic suction that draws free water from the surrounding into the freezing zone where ice lenses are formed. This zone is often referred to as the *frozen fringe*. The ice lenses continue to grow until the freezing front with its trailing frozen fringe penetrate the ground

further, causing new ice lenses to nucleate and grow, while retarding the growth of ice lenses left behind. Alternatively, the growth of ice lenses can be reduced by exhaustion of free water available in the proximity of the freezing front (e.g., in unsaturated soils). As the ice lenses grow, they generate the heaving pressure; and the ground surface heaves when the heaving pressure is large enough to overcome the overburden pressure.

Frost heave is one of the major causes of damage to infrastructure, such as earth retaining walls, in cold regions. The heaving pressure in addition to the usual earth pressure acting on the structures may cause unacceptable deformation, or even structural failure, if it is not properly considered in design. Field measurements of frost heave and heaving pressures on earth retaining structures can be found in literature. McRostie and Schriever (1967) reported that considerable frost pressures averaging 300 to 600 psf (or 14.4 to 28.7 kPa) were observed on a tied-back excavation bracing system. The measured horizontal wall movement was about 0.07 ft (or 21 mm), which appeared to be smaller than expected. Morgenstern and Sego (1981) reported seasonal stress changes in an anchored tie-back wall; the peak anchor load was observed in mid-February, when the temperatures dropped below freezing. Sandegren et al. (1972) presented a case history where an anchored sheet pile wall failed when it was exposed to frost action. It was reported that, in winter time, the anchor load had exceeded 50% to 100% the designed value, and resulted in the wall failure.

Although case histories and field measurements are available, there has been limited attempt toward modeling and prediction of frost heave for earth retaining structures. This is mainly due to complexity of the problem that involves the interrelated processes of heat transfer, water transport, stress, and deformation in a multi-phase porous material (i.e., soil). This paper presents a numerical approach that can be applied to simulate the freezing process and associated frost heave in soil behind an earth retaining structure. In this approach, soil is treated as a mixture of solid particles, unfrozen water, and ices. The conduction through soil components is considered as the primary mode of heat transfer. The expansion of soil due to formation and growth of ice lenses is described mathematically using the Porosity Rate Function, where the porosity rate (i.e., change of soil porosity over time) is expressed as a function of the soil temperature, temperature gradient, and stress. The heat transfer and frost heave models have been incorporated into a finite element program for numerical simulations.

The next section of the paper describes the details of the numerical modeling, including the heat transfer model, the frost heave model, the finite element analysis, and the model calibration and validation. A case is then presented where a concrete cantilever earth retaining wall was simulated under freezing conditions. The use of thermal insulation to control the wall deformation was also evaluated. The paper concludes with discussion of the modeling and the simulation results.

NUMERICAL MODELING

Heat Transfer Model

There are three basic modes of heat transfer: conduction, convection, and

radiation. The heat transfer by convection or radiation is usually considered to be negligible in soil (Farouki, 1981). The conduction through soil components is considered as the primary mode of heat transfer. The governing differential equation for the heat transfer is:

$$C\frac{\partial T}{\partial t} - L\rho_i \frac{\partial \theta_i}{\partial t} - \nabla(\lambda \nabla T) = 0$$

where, T is the temperature (°C), t is the time (s), C is the volumetric heat capacity of the soil mixture (J/m³·°C), ρ_i is the density of ice (kg/m³), θ_i is the volumetric fraction of ice (m³/m³), L is the latent heat of fusion per unit mass of water (approximately 3.33×10^5 J/kg), ∇ is the gradient operator, and λ is the thermal conductivity of the soil mixture (W/m·°C). The volumetric heat capacity and thermal conductivity of the soil mixture can be expressed as follows:

$$C = \rho_w c_w \theta_w + \rho_i c_i \theta_i + \rho_s c_s \theta_s$$

$$\lambda = \lambda_s^{\theta_s} \lambda_w^{\theta_w} \lambda_i^{\theta_i}$$

where, c is the mass heat capacity (J/kg·°C) and subscripts w, i and s denote water, ice, and soil particles, respectively. A 3-parameter function is used to describe the unfrozen water content in soil (Michalowski, 1993):

$$w = w^* + (\overline{w} - w^*)e^{a(T-T_0)}$$

where, $\overline{w}$ is the unfrozen water content at the freezing temperature T_0, w^* is the residual unfrozen water content at a low reference temperature, and parameter a describes the rate of decay.

The heat transfer model presented above takes into account the following key factors related to the soil freezing process: (1) the release or absorption of latent heat during the phase change of water; (2) the change of soil thermal properties as a result of the change in soil composition; and (3) the existence and effect of unfrozen water at temperatures below freezing (i.e., the supercooled water).

Frost Heave Model

The frost heave model is based on the Porosity Rate Function (Michalowski and Zhu, 2006 and Zhu, 2006):

$$\dot{n} = \dot{n}_m \left(\frac{T-T_0}{T_m}\right)^2 \cdot e^{1-\left(\frac{T-T_0}{T_m}\right)^2} \cdot \frac{\left|\frac{\partial T}{\partial l}\right|}{g_T} \cdot e^{-\frac{|\overline{\sigma}_{kk}|}{\zeta}}$$

where, $\dot{n}_m$ is the maximum porosity rate for a given soil, T_m is the temperature (°C) at which this maximum porosity rate occurs, T (°C) is the temperature, T_0 (°C) is the freezing point of water, $\partial T/\partial l$ is the temperature gradient, g_T is the temperature gradient corresponding to $\dot{n}_m$, $\overline{\sigma}_{kk}$ is the first invariant of the total stress tensor, and ζ is a material parameter.

To model the anisotropic growth of ice lenses, the porosity growth tensor $\dot{n}_{ij}$ is introduced:

$$\dot{n}_{ij} = \dot{n}\alpha_{ij}$$

where,

$$\alpha_{ij} = \begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} = \begin{vmatrix} \xi & 0 & 0 \\ 0 & (1-\xi)/2 & 0 \\ 0 & 0 & (1-\xi)/2 \end{vmatrix}$$

The dimensionless quantity ξ ranges between 0.33 (i.e., isotropic growth) and 1.0 (i.e., one-dimensional growth).

The deformation of soil is assumed to be elastic. The strain increment consists of both the elastic strain increment and the growth increment due to the change in porosity:

$$d\varepsilon_{11} = \frac{1}{E}\left[d\sigma_{11} - \mu(d\sigma_{22} + d\sigma_{33})\right] + \xi\dot{n}dt$$

$$d\varepsilon_{22} = \frac{1}{E}\left[d\sigma_{22} - \mu(d\sigma_{11} + d\sigma_{33})\right] + \frac{1}{2}(1-\xi)\dot{n}dt$$

$$d\varepsilon_{33} = \frac{1}{E}\left[d\sigma_{33} - \mu(d\sigma_{11} + d\sigma_{22})\right] + \frac{1}{2}(1-\xi)\dot{n}dt$$

$$d\gamma_{12} = \frac{\tau_{12}}{G}, \quad d\gamma_{23} = \frac{\tau_{23}}{G}, \quad d\gamma_{31} = \frac{\tau_{31}}{G}$$

where, E and μ are Young's modulus and Poisson's ratio, respectively, and the shear modulus G is equal to $E/2(1 + \mu)$.

The Porosity Rate Function model presented above describes the growth of ice lenses as the change in soil porosity over time and thus, models the global response of the freezing soil. This model is distinctly different from other models, which focus on the growth of individual ice lenses (e.g., the rigid ice model of Miller, 1978), and is easier to implement using standard numerical techniques.

Finite Element Analysis

The heat transfer and frost heave models described above have been incorporated into the commercial finite element program ABAQUS through its user subroutines UMATHT and UEXPAN (see Figure 1). The two subroutines are called at all integration points of the finite elements at each time increment. The analysis is a fully coupled thermal-displacement analysis, i.e., the temperature and stress distributions at the end of each time increment affect the porosity change of soil in the next time increment; on the other hand, the porosity change and the volumetric composition of soil components affect the thermal properties of soil and thus the temperature distribution.
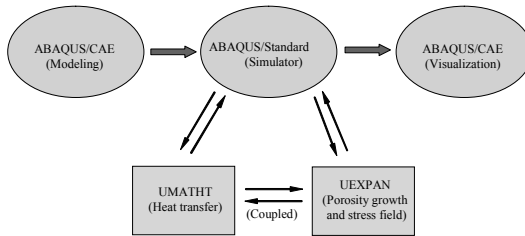
Figure 1. Illustration of finite element analysis

Model Calibration and Validation

The Porosity Rate Function model for frost heave was calibrated and validated using the results of laboratory soil freezing tests by Fukuda et al. (1997) (Michalowski and Zhu, 2006 and Zhu, 2006). In the model calibration process, the three parameters in the porosity rate function, $\dot{n}_m$, T_m, and ζ, were obtained based on the results of a series of ramped freezing tests conducted under different overburden pressures. The calibration involved the curve fitting process to find the best values of the model parameters in order to match the calculated results with the experimental results. In the model validation process, the calibrated model was applied to simulate other freezing tests conducted on the same material. The simulation results were found to agree well with the experimental results.

Evaluation and comparison of the Porosity Rate Function model with two other frost heave models, i.e., the hydrodynamic model FROST (Guymon et al., 1993) and the rigid ice model PC-Heave (Sheng, 1994), was carried out by Henry et al. (2005). It was found that the Porosity Rate Function model and the PC-Heave procedure both accurately simulated the laboratory soil freezing tests. More recently, the model presented in this paper was applied to simulate the artificial ground freezing associated with subway tunnel construction. The ground heaving predicted by the model was reported to be within 10% of the field measurement (Chiu, 2010).

SIMULATION OF A CANTILEVER RETAINING WALL

A typical concrete cantilever earth retaining wall was simulated using the numerical approach described in the previous section. The total height of the wall is 6.5 m with 1.5 m buried under the ground. The base of the wall is 4 m wide. A 0.1 m thick thermal insulation layer is included behind the wall. The model is discretized into approximately 3,000 4-node plane strain thermally coupled quadrilateral elements (element type CPE4T in ABAQUS). The two vertical external side boundaries are fixed in the horizontal direction, while the bottom boundary is fixed in both horizontal and vertical directions. Heat flow is not allowed at the two vertical external side boundaries.

The initial temperature distribution is shown in Figure 2a. This is the steady state distribution associated with constant temperatures at the top and bottom boundaries as indicated in the figure. The thermal boundary conditions are defined as

follows: at time t = 0, the temperature along the top external boundary drops from 1°C to -5°C and is kept at this level for 3 months, while the temperature at the base is kept steady at 3°C. The initial vertical stress distribution in the soil under the self-weight is shown in Figure 2b. The backfill is considered to be perfectly bonded to the wall causing tensile stress near the tip and toe of the wall.

(a)   (b)
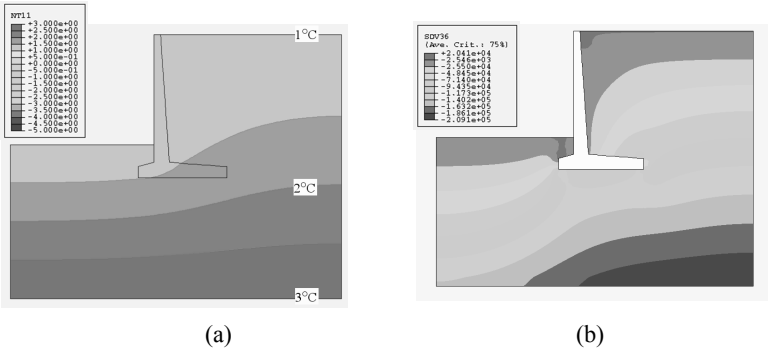
Figure 2. Initial temperature (left) and vertical stress (right) distributions

(a)   (b)
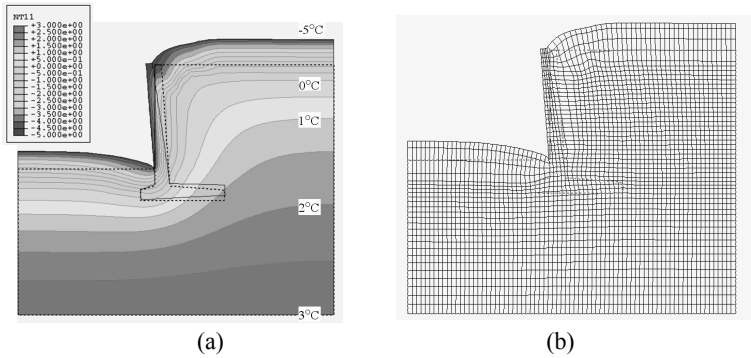
Figure 3. Temperature distribution and deformation (no thermal insulation)

Table 1. Summary of Material Properties.

	Density ρ (kg/m^3)	Mass heat capacity c (J/kg·°C)	Thermal conductivity λ (W/m·°C)	Young's modulus E (MPa)	Poisson's ratio μ
Concrete	2242	970	1.20	2×10^4	0.38
Thermal insulation	50	2000	0.03	10	0.30
Soil particle (Clay)	2620	900	1.95	11.2 for unfrozen soil, 13.75 $|T|^{1.18}$ for frozen soil below -1°C, and linear interpolation between 0 and -1°C	0.30 for unfrozen and frozen soils
Water	1000	4180	0.56		
ice	917	2100	2.24		

Note: The properties were compiled according the values found in literature (Ladanyi and Shen, 1993, Williams and Smith, 1989, and Selvadurai et al., 1999).

The parameters for the Porosity Rate Function model were obtained from the frost heave tests performed by Fukuda et al. (1997) on a frost-susceptible silty clay material. Based on the test results, the following values were interpreted: $\dot{n}_m = 6.02 \times 10^{-5}$ /s at $g_T = 100\,°C/m$ (or $\dot{n}_m / g_T = 6.02 \times 10^{-7}$ m/°C·s), $T_m = -0.87\,°C$, and $\zeta = 0.6$ MPa. Parameter ξ that governs the anisotropy of the ice growth was selected to be 0.9 because ice lenses grow predominantly in the direction of heat flow. The initial porosity of the soil was considered to be 0.43. The thermal and elastic properties of the concrete, insulation, and soil components are summarized in Table 1.

Simulation was first conducted for the retaining wall without the thermal insulation. The calculated temperature distribution and wall and ground deformation are shown in Figure 3. The simulation results show that as the frost front penetrates into the ground, the soil behind the retaining wall expands in both horizontal and vertical directions. The wall tilts gradually as the soil keeps pushing it. The horizontal wall displacement at the tip was calculated to be 0.21 m. In order to evaluate the effect of insulation on controlling frost heave, the wall was simulated again with the thermal insulation layer behind the wall. The horizontal wall displacement at the tip was calculated to be 0.12 m, which is a reduction in the wall movement by 42%.

SUMMARY

A numerical approach was developed to model and simulate the soil freezing and frost heave, which incorporates the heat transfer model for soil, the Porosity Rate Function model for frost heave, and the finite element method. The Porosity Rate Function model is distinctly different from some of the other frost heave models. It is based on the concept of the porosity growth described in terms of the continuum. With a formulation that is consistent with continuum mechanics, it becomes possible to generalize the model and use standard numerical techniques to solve boundary value problems. The numerical approach was used to simulate a typical earth retaining structure. The simulation results appear to be quite realistic. The additional simulation considering wall insulation demonstrates that the wall deformation can be effectively controlled by installing a thermal insulation layer behind the wall. The numerical approach described in the paper can be used by engineers or researchers to improve the design of earth retaining structures in cold regions.

ACKNOWLEDGMENT

REFERENCES

Chiu, Shih-Ming. (2010). *Comparative analysis of frost heave model with artificial ground freezing.* Master Thesis, National Central University.
Henry, K.S., Zhu, M., and Michalowski, R.L. (2005). "*Evaluation of three frost heave*

models." *Proceedings of 7th International Conference on the Bearing Capacity of Roads, Railways and Airfields (BCRA'05)*, Trondheim, Norway, June 2005

Guymon, G.L., Berg, R.L. and Hromadka, T.V. (1993). "Mathematical model of frost heave and thaw settlement in pavements." *CRREL Report, 93-2*, US Army Cold Regions Research and Engineering Laboratory, Hanover, MH.

Farouki, O.T. (1981). "Thermal properties of soils." *CRREL Monograph 81-1*. U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, NH.

Freitag, D.R. and McFadden, T.T. (1996). *Introduction to cold regions engineering*, American Society of Civil Engineers.

Fukuda, M., Kim H. and Kim Y. (1997). "Preliminary results of frost heave experiments using standard test sample provided by TC8." *Proceedings of the International Symposium on Ground Freezing and Frost Action in Soils*, Luleå, Sweden, 25-30.

Ladanyi, B. and Shen, M. (1993). "Freezing pressure development on a buried chilled pipeline." *Proceedings of the 2nd International Symposium on Frost in Geotechnical Engineering*, Anchorage, AK, 23-33.

McRostie, G.C. and Schriever, W.R. (1967). "Frost pressures in tie-back system at national arts centre excavation." *Engineering Journal*, Vol. 50, No. 2, 17-21.

Michalowski, R.L. (1993). "A constitutive model of saturated soils for frost heave simulations." *Cold Regions Science and Technology*, Vol. 22, No. 1, 47-63.

Michalowski, R. L. and Zhu, M. (2006). "Frost heave modelling using porosity rate function." *International Journal for Numerical and Analytical Methods in Geomechanics*, 30(8), 703-722.

Miller, R. D. (1978). "Frost heaving in non-colloidal soils." *Proceedings of the Third International Conference on Permafrost*. Edmonton, 707-713.

Morgenstern, N. R. and Sego, D. (1981). "Performance of temporary tie-backs under winter conditions." *Canadian Geotechnical Journal*, Vol. 18, No. 4, 566-572.

Mutou, Y., Watanabe, K., Ishizaki, T., and Mizoguchi, M. (1998). "Microscopic observation of ice lensing and frost heave in glass beads." *Permafrost – Seventh International Conference*, Yellowknife (Canada), Collection Nordicana, No 55, 783-787.

Sandegren, E., Sahlstrom, P. O., and Stille, H. (1972). "Behaviour of anchored sheet pile exposed to frost action." *Proceedings of the 5th European Conference on Soil Mechanics and Foundation Engineering*, Madrid, Vol. 1, 285-291.

Sheng, D. (1994). *Thermodynamics of freezing soils: theory and application*. Doctoral Thesis. Luleå University of Technology, Sweden.

Selvadurai A.P.S., Hu J., and Konuk I. (1999). "Computational modeling of frost heave induced soil-pipeline interaction: I. Modeling of frost heave." *Cold Regions Science and Technology*, Vol. 29, pp. 215-228.

Taber, S. (1930). "The Mechanics of Frost Heaving." *The Journal of Geology*, Vol. 38, No. 4 (May - Jun., 1930), 303-317.

Williams, P. J. and Smith, M. W. (1989). *The Frozen Earth, Fundamentals of Geocryology*. Alden Press, Oxford, pp. 90.

Zhu, M. (2006). *Modeling and simulation of frost heave in frost-susceptible soils*. Ph.D. thesis, University of Michigan, Ann Arbor.

Granular Element Method for Computational Particle Mechanics

Keng-Wit Lim and José E. Andrade

Division of Engineering and Applied Sciences,
California Institute of Technology, MC 104-44, Pasadena, CA 91125;
PH 1-626-395-4141; email: kengwit@caltech.edu, jandrade@caltech.edu

ABSTRACT

We present and review the Granular Element Method (GEM), which is a discrete mechanics simulation technique that uses Non-Uniform Rational Basis Splines (NURBS) for capturing particle morphology and inferring grain-scale kinematics and forces at computer-tomographic resolutions. The target application of GEM is in analyzing granular systems in advanced experiments utilizing visualization tools such as X-ray computer tomography experiments (XRCT). In this context, GEM may help in facilitating the transition from binary data from XRCT to discrete models, as well as lead to more realistic predictions at the grain scale.

INTRODUCTION

The motivation behind the Granular Element Method (Andrade et al. 2012a) originates from the need to probe granular assemblies at the grain scale using imaging experiments (Andrade et al. 2012b, Vlahinich et al. under review), in particular X-ray computed tomography (XRCT). At this scale, grain kinematics becomes an important micromechanical measure to understand macroscopic behaviour and current shape representation techniques such as clustering or clumping does not provide the required fidelity (Andrade et al. 2012a). Some features that are unique to GEM include the ability of NURBS to capture key particle morphology features, namely sphericity and roundness, without further numerical treatment, and the availability of local curvature information. The NURBS basis can also serve as a platform for integration with isogeometric analysis methodologies (Hughes et al. 2005). With these enhancements, discrete NURBS models derived from binary image data are used on-the-fly in discrete computations. We note that GEM uses the same explicit time-integration procedure as classical DEM (Cundall and Strack 1979).

In this paper, we focus only on the salient ingredients required for the implementation of GEM, namely NURBS and the 3D contact algorithm. We also briefly mention the development of a contact dynamics approach to GEM, in which the system is solved implicitly accounting for non-penetrating constraints. Two

numerical examples are presented to showcase the capabilities of GEM. We then conclude and discuss current limitations and avenues for improving GEM.

NON-UNIFORM RATIONAL BASIS-SPLINES (NURBS)
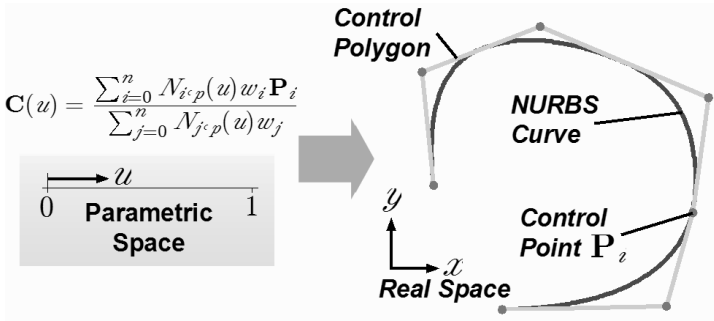
Figure 1. Schematic of NURBS curve

We define grain shapes in 2D and 3D using NURBS curves and surfaces, respectively. The NURBS curve is defined parametrically as (Piegl and Tiller 1997):

$$C(u) = \frac{\sum_{i=0}^{n} N_{i,p}(u) w_i P_i}{\sum_{j=0}^{n} N_{j,p}(u) w_j} \tag{1}$$

where P_i are the control points, u is the curve parameter or knot value, $N_{i,p}$ is the i-th B-Spline basis function of degree p defined recursively as follows:

$$N_{i,0} = \begin{cases} 1 & \textit{if } u \in [u_i, u_{i+1}) \\ 0 & \textit{otherwise} \end{cases} \quad , \quad N_{i,p} = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+p}} N_{i+1,p-1}(u) \tag{2}$$

and w_i are the weights, which provides an additional degree of freedom for geometry manipulation. The control points need not lie on the curve itself. Rather, control points define a control polygon that acts as a scaffold that controls the geometry (see Figure 1). By taking a tensor product of the curve case, one obtains a NURBS surface patch, described parametrically (in s, t) as:

$$S(s,t) = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} w_{ij} N_{i,p}(s) N_{j,p}(t) P_{ij}}{\sum_{i=0}^{m} \sum_{j=0}^{n} w_{ij} N_{i,p}(s) N_{j,p}(t)} \tag{3}$$

where P_{ij} are the control points, w_{ij} are the weights, and $N_{i,p}(s)$ and $N_{j,p}(t)$ are the B-Splines univariate basis functions of degree p, as in Eq. (2). For most cases, grain surfaces are adequately described as a bicubic patch $p = 3$.

Several NURBS modeling tools are already available to aid the construction of NURBS grains (for example, Rhino3D). For example, geometrical quantities such as centroid, volume, moments of inertia and product moments for each grain can be numerically evaluated using Rhino3D. This ability to integrate tightly with and quickly access geometrical information from standard NURBS CAD tools highlights GEM's flexibility during model generation.

CONTACT ALGORITHM

We restrict the discussion here to the 3D contact algorithm, which is a nontrivial extension of the 2D case described in (Andrade et al. 2012a). Referring to Figure 2, an intersection-based contact algorithm is devised in which potentially intersecting NURBS surfaces in real space are adaptively subdivided in both parametric directions. The subdivision procedure is repeated until a pre-specified tolerance for surface flatness, in parametric space, is reached. At this level, the isolated surface is discretized into triangles to facilitate surface-to-surface intersection and approximate the intersection loop. To determine the contact point on each surface,
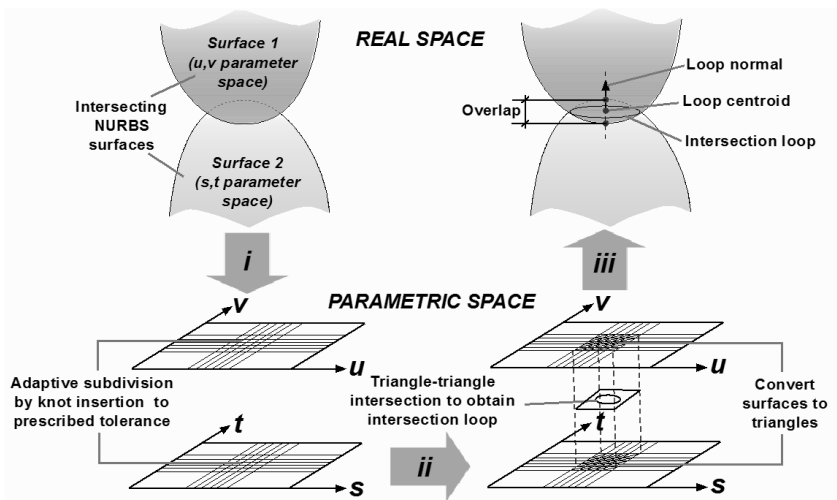
Figure 2. Contact algorithm in 3D GEM (Lim and Andrade, under review)

we employ a curve-surface intersection algorithm, with the curve passing through the loop centroid and in the direction of the loop normal. With the intersection loop normal and contact point on each surface determined, the overlap is calculated and the normal and tangential contact forces are calculated as in classical DEM. Collision detection algorithms for polyhedral-based methods can be adapted to GEM, making its computational complexity comparable with polyhedral-based methods.

A CONTACT DYNAMICS APPROACH TO GEM

The so-called non-smooth contact dynamics (CD) method (Moreau 1994, Jean 1999) is an alternative to the classical DEM. CD targets granular systems comprising of rigid or highly stiff particles, conditions that result in the poor performance of DEM due to infinitesimally small time steps. The most prominent feature of CD, in contrast to that of classical DEM, is that the contact forces and particles are considered perfectly rigid and the contact forces are determined as those that exactly prevent inter-particle penetration. We note that particle elasticity can be easily accounted for through a regularization scheme (Krabbenhoft et al. 2012a). For simplicity, we present here only the case of perfectly rigid particles. CD employs an implicit solution procedure in which both the constraint and contact forces are simultaneously solved for and larger time steps can be taken. We adopt a recent contact dynamics formulation (Krabbenhoft et al. 2012b), which casts the discrete problem into a quadratic programming problem of the form:

$$\text{Minimize} \quad \frac{1}{2} \boldsymbol{r}^T \boldsymbol{A}\boldsymbol{r} + \frac{1}{2} \boldsymbol{t}^T \boldsymbol{B}\boldsymbol{t} + \boldsymbol{g_0}^T \boldsymbol{p}$$

$$\text{subject to} \quad \boldsymbol{r} + \boldsymbol{N_0}\boldsymbol{p} + \hat{\boldsymbol{N_0}}\boldsymbol{q} = \bar{\boldsymbol{f_0}}, \quad \boldsymbol{t} - \boldsymbol{R_0^q}\boldsymbol{p} - \boldsymbol{R_0^p}\boldsymbol{q} = \bar{\boldsymbol{m_0}}, \quad \|\boldsymbol{q}\| - \mu \boldsymbol{p} \le \boldsymbol{0}, \quad \boldsymbol{p} \ge \boldsymbol{0}$$

where $\boldsymbol{q}$, $\boldsymbol{p}$ are the contact normal and shear forces, $\boldsymbol{r}$, $\boldsymbol{t}$ are the dynamic forces and torques. $\boldsymbol{N_0}$, $\hat{\boldsymbol{N_0}}$ describe the contact topology and $\boldsymbol{R_0^q}$, $\boldsymbol{R_0^p}$ contains the moment arms associated with the contact forces. The mass and moment of inertia terms are embedded in $\boldsymbol{A}$ and $\boldsymbol{B}$. The external forces and moments are denoted by $\bar{\boldsymbol{f_0}}$ and $\bar{\boldsymbol{m_0}}$, respectively. The particle interpenetrations, to be corrected in the subsequent time, are embedded in $\boldsymbol{g_0}$. In the static limit case, the quadratic terms associated with $\boldsymbol{A}$ and $\boldsymbol{B}$ drop out and the above reduces to a linear programming problem. The above problem can be solved using standard optimization techniques. We are currently developing a contact dynamics approach to GEM, which combines particle shape flexibility, properties of implicit time-integration discretization (such as large time steps) and non-penetrating constraints, as well as reduction to a static formulation in the limit of an infinite time step (Lim et al., in preparation). These features are particularly desirable in modeling rigid or highly stiff particles, especially under quasi-static conditions.

NUMERICAL EXAMPLES

In this section, we present two numerical examples to highlight the capabilities of GEM. Verification examples can be found in (Andrade et al. 2012ab). Integration of GEM with experimental data from XRCT is currently underway and will be reported in the near future.

Mixed-Boundary Test. A mixed-boundary test is performed on a 3D assembly with 240 GEM particles (Lim and Andrade, under review). This problem is solved using an explicit method with dynamic relaxation. The interparticle friction

angle is 15 degrees and the friction angle between the walls and particles is zero. The assembly is initially subjected to a specified confinement pressure. Two confinement pressures are considered: 5 and 10 kilounits. Subsequently, a vertical pressure of approximately 5 and 10 kilounits, respectively, is maintained on the top wall. Simultaneously, the side walls in the x-direction are rotated about y-axis at the base to an angle of approximately 0.15 radians (8.6 degrees); the walls in the y-direction are held fixed i.e. plane strain shear. The deformed configuration of the assembly with initial confinement of 5 kilounits after the shearing process is shown in Figure 3 together with the plots of stress ratio q/p and volumetric strain ε_v against the shear angle. Both stress ratios peak near the phase transformation point i.e. when the volumetric strain starts to change direction (the lack of a clear phase transformation is due to the dense packing of the assembly). In both cases, the stress ratio and volumetric strain becomes approximately constant at later stages, indicating critical state behavior. For two samples with approximately the same void ratio, the one with the lower confinement exhibits larger stress ratio and dilation. These characteristics are the same as those observed in classical soil mechanics behavior. We note that the volumetric strain here is somewhat small due to the small assembly size. Some oscillations in the solution are present due to the dynamic nature of the explicit method and a servomechanism used to control the top wall pressure based on the particle-wall overlap, which changes as the assembly is sheared.
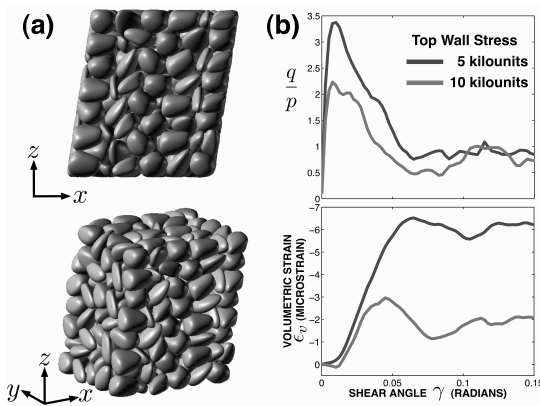
Figure 3. Mixed-boundary test (Lim and Andrade, under review): (a) assembly configurations after shear (initial confinement of 5 kilounits case), (b) Plot of stress ratio and volumetric strain against shear angle.

Biaxial Compression. A 2D biaxial test is carried out for a rectangular assembly with 960 particles as shown in Figure 2. Also shown in the figure is another assembly constructed using disks, which is used for comparison. Both assemblies are densely packed, with an initial porosity of 0.175. The upper wall is moved downwards while the pressure on the right wall is maintained at 125 units of pressure.

A total of 300 steps are used to impose a total axial strain of approximately 0.15. The left and bottom walls are stationary and all walls are frictionless. Additionally, to see the effects of interparticle friction, three friction coeffcients are considered: $\mu = 0.354, 0.577, 0.839$. This problem is solved in the static limit using the contact dynamics approach (Lim et. al., in preparation). The test results are shown in Figure 4, where we see that both the peak strength and dilatancy (as measured by the initial slope in the volumetric strain plot i.e. volumetric change near the peak strength) in the GEM assembly are higher than those in the disc assembly. These observations are consistent with the increase in angularity of the particles.
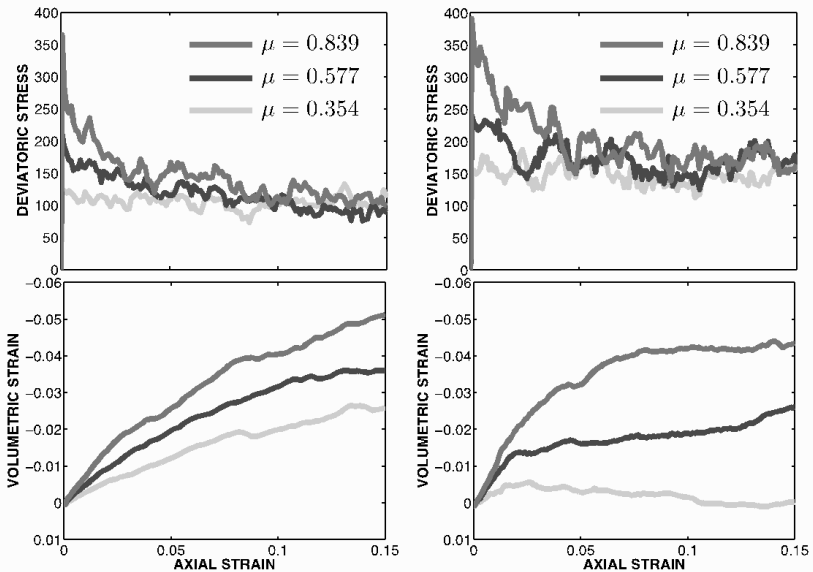
Figure 4. Biaxial test (Lim et. al., in preparation): (a) disk assembly (b) GEM assembly.

CONCLUSION

We have presented and reviewed the development of the Granular Element Method (GEM), which targets the discrete analysis of granular systems in advanced experiments utilizing visualization tools such XRCT. Shape representation at the grain-scale is achieved using NURBS. Grain geometrical information is already available for natural granular materials such as sands (Andrade et al. 2011). Leveraging on existing NURBS geometric algorithms, we presented a 3D algorithm to perform local contact detection and contact force calculation. Numerical examples are presented to show the functionality GEM with complicated grain geometries. We are currently interfacing GEM with binary image data from XRCT and a quantitative comparison of particle kinematics will be reported in the near future.

We conclude by highlighting two areas that are required to improve the applicability and performance of GEM. The first concerns the interface between experimental data and computations. In this aspect, there is a need for a fully automatic and direct fitting and optimization of grain shapes and multi-patch design of parametric surfaces. The second concerns the contact algorithm, which is the current computational bottleneck of GEM. Here, the difficulty stems from the nonlinear nature of parametric surfaces, which makes the contact algorithm an iterative process. Further research is required to identify more efficient strategies to enhance the subdivision and intersection procedures leading to the computation of contact forces.

REFERENCES

Andrade, J. E., Vlahinich, I., Lim, K.-W., and Jerves, A. (2012a). Multiscale 'tomography-to-simulation' framework for granular matter: the road ahead, *Géotechnique Letters*, Vol. 2, Issue July-September.

Andrade, J. E., Lim, K.-W., Avila, C. F., and Vlahinich, I. (2012b). Granular Element Method for Computational Particle Mechanics, *Computer Methods in Applied Mechanics and Engineering*, 241-244:262-274.

Andrade, J. E., Avila, C. F., Hall, S. A., Lenoir, N., and Viggiani, G. (2011). Multiscale modeling and characterization of granular matter: from grain kinematics to continuum mechanics. *Journal of the Mechanics and Physics of Solids*, 59:237–250.

Cundall, P. A. and Strack, O. D. L. (1979) A discrete numerical model for granular assemblies. Geotechnique, 29:47-65.

Hughes, T. J. R., Cottrell, J. A., Bazilevs, Y. (2005). Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement, *Computer Methods in Applied Mechanics and Engineering*, 194(39-41):4135-4195.

Jean, M. (1999) The non-smooth contact dynamics method. *Computer Methods in Applied mechanics and Engineering* 177(3-4):235-257.

Krabbenhoft, K., Lyamin A.V., Huang J., and Silva, M. V. (2012a) Granular contact dynamics using mathematical programming methods. *Computers and Geotechnics*, 241-244:262–274.

Krabbenhoft, K., Lyamin A.V., Huang J., and Silva, M. V. (2012b) Granular contact dynamics with particle elasticity. *Granular Matter*, 14, 607-619.

Lim, K.-W., Krabbenhoft, K., and Andrade, J. E. (in preparation). A contact dynamics approach to the Granular Element Method.

Lim, K.-W. and Andrade, J. E. (under review). Granular Element Method (GEM) for three-dimensional discrete element calculations, *International Journal of Analytical and Numerical Method in Geomechanics*.

Moreau, J.J. (1994). Some numerical methods in multibody dynamics: application to granular materials. *Eur. J. Mech. A Solids*, 13, 93–114.

Piegl, L. and Tiller, W. (1997). The NURBS book (2nd ed.). Springer-Verlag New York, Inc., New York, NY, USA.

Vlahinic, I., Andrade, J. E., Ando, E., and Viggiani, G. (under review) A method for inferring accurate grain shape and fabric in discrete media using full-fidelity tomographic images. *Granular Matter*.

Requirements to enhance the decision-making process for tunnel construction by Virtual Design and Construction (VDC)

Jung In Kim[1] and Martin Fischer[1]

[1]Center for Integrated Facility Engineering (CIFE), Dept. of Civil and Environmental Engineering, Stanford University, Y2E2 Building, 473 Via Ortega, Room 292, Stanford, CA 94305, United States of America; Phone: +1-650-723-2300; Fax: +1-650-723-4806; email: jikim07@stanford.edu

Abstract

This paper discusses requirements to enhance the current decision-making process for tunnel construction, which involves significant uncertainties in ground conditions. In current practice, project managers determine resource-loaded construction schedules based on one specific set of ground conditions. In addition, they do not update ground conditions for unexcavated sections with information from excavated sections. This decision-making process often causes cost overruns, schedule delays, and disputes. To mitigate these problems, managers must make timely and robust decisions about construction schedules, during both the preconstruction and construction phases, in a transparent and integrated manner. To this end, we applied a 4D model-based decision-support system, the Development Strategy Simulator (DSS), to a U.S. tunnel project. Our analysis showed that the DSS was able to simultaneously analyze and represent the construction schedule and the resulting resources. However, we argue that it is still necessary to develop a progress-adaptive stochastic earth model, as well as a formal process for the automatic formulation and evaluation of multiple alternative schedules.

Introduction

Tunnel construction involves significant uncertainties in ground conditions that affect both tunnel design (e.g., support systems) and excavation productivities. Although geo-technical investigation is performed before the construction phase, a certain degree of deviation often occurs between predicted and actual ground conditions. For this reason, project managers for tunnel construction projects develop resource-loaded construction schedules and then modify the schedules as construction progresses.

However, since the current manual and experience-based decision-making process for tunnel construction schedules is not fast, integrated, transparent, or robust, project managers often encounter difficulty completing their projects on time and within budget. Such difficulties often lead to disputes among the stakeholders. For example, in current practice, the managers determine their construction schedules based on one specific set of ground conditions. In addition, when they modify the construction schedules during the construction phase, they do not update ground conditions for unexcavated sections of the tunnel with given information about ground conditions from the excavated sections. Moreover, they cannot present their

schedule and resulting resources to other stakeholders (e.g., owner, subcontractors, and material suppliers) in a timely, transparent, and integrated manner.

To mitigate those problems, our research team investigated a recently-constructed tunnel project and identified the challenges to enhance the decision-making process for tunnel construction schedules. Next, the team investigated previous studies about how to overcome these challenges. The team then applied a 4D model-based decision-support system, the Development Strategy Simulator (DSS), to a U.S. tunnel construction project. The DSS allows managers to simultaneously visualize the construction schedule and represent consequent metrics (i.e., particular resources required). Based on the results, the team discussed the remaining challenges and the requirements to overcome those challenges.

Motivating case

This section describes problems encountered in tunnel construction by using a case study and challenges identified to make better decisions on the tunnel construction schedules. The case study highlights significant uncertainties in ground conditions that affect both support systems (e.g., shotcrete and rockbolts) and excavation productivities for tunnel construction. Panthi et al. (2007) compared the predicted and actual rock mass conditions for four recently-constructed hydro-tunnels in Nepal Himalaya: the Khimti 1, the Kaligandaki "A", the Modi Khola, and the Middle Marsyangdi headrace tunnels. In addition, they analyzed variations in quantities between planned and actual tunnel support systems.
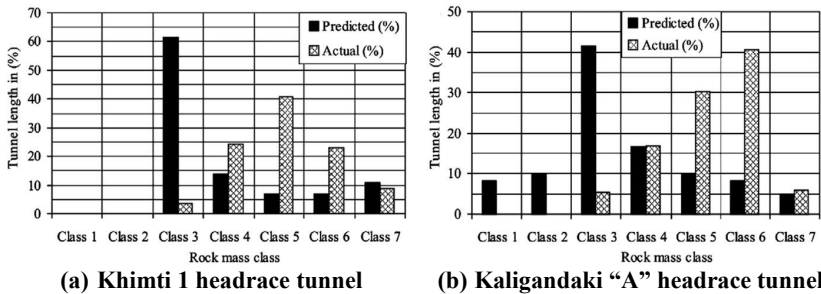
(a) Khimti 1 headrace tunnel　　　(b) Kaligandaki "A" headrace tunnel
Figure 1. Predicted and actual rock mass conditions (Panthi et al., 2007)

As the classification of rock mass conditions is essential for estimating required tunnel rock support and construction time, the rock mass conditions for the four tunnels were classified based on the Q and RMR systems (Barton 1995; Bieniawski 1989): Class 1 is very good to excellent, while Class 7 is exceptionally poor. Considerable differences existed between predicted and actual rock mass conditions for the four tunnels. In particular, the discrepancies for the Khimti 1 and the Kaligandaki "A" headrace tunnels were found to be high. Although Class 3 was dominant in the predicted condition for both tunnels, Class 5 and Class 6 were dominant in the actual conditions for the Khimti 1 and the Kaligandaki "A" tunnels,

respectively (see Figure 1 (a) & (b)). As a result of the large discrepancy between predicted and actual rock mass conditions, the differences between predicted and actual tunnel support systems were also large for these two projects (see Figure 2).
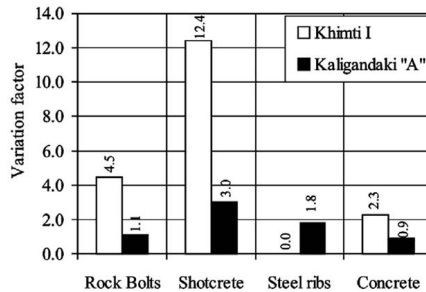
Figure 2. Variation factors in tunnel support quantities between predicted and actual for two tunnels (Panthi et al., 2007)

These large differences in ground conditions and consequent support systems resulted in disputes among the stakeholders, especially because the general contractor (GC) had spent more money on the project to accelerate the project schedule. While excavating the tunnel, the GC found significant differences between the actual and predicted ground conditions. As the actual ground conditions were much worse than those predicted in the preconstruction phase, the GC needed to accelerate the project schedule because of the design-build contract. For this reason, the GC excavated an additional adit to excavate the main tunnel on two more faces. In addition to spending more money to excavate the additional adit, the GC was forced to increase expenditure to recover the delayed schedule caused by differing ground conditions because the GC was unable to make a decision about the additional adit in a timely manner. This unnecessary cost overrun resulted in disputes between the owner and the GC and led to additional money spent for litigation.

To mitigate such problems, our team investigated the existing decision-making process for tunnel construction and identified four main challenges. First, the project management team used one specific set of ground conditions that geotechnical engineers provided for the management team. Second, the management team was unable to consistently update the ground conditions for unexcavated sections by using additional information about ground conditions as the construction progressed, even though the project management team updated ground conditions for excavated sections. Third, the management team manually, subjectively, and separately analyzed multiple performance metrics (e.g., cost, schedule, and resources required) with limited alternative construction schedules. Finally, the management team had difficulty delivering the modified schedule and consequent resources to other stakeholders transparently and quickly.

Points of departure

This section reviews previous studies about how to overcome the challenges of decision making for the tunnel construction schedules. According to the principle of lean construction, project managers should deliver transparent and timely information to the stakeholders to reduce unnecessary costs. In addition, a probabilistic approach is required to mitigate risks caused by uncertain ground conditions. Finally, Virtual Design and Construction (VDC) methodology is explored to improve the decision-making process for tunnel construction.

Lean Construction

The principle of lean construction stems from lean production, a concept that aims to systematically reduce waste (e.g., unnecessary cost), simplify production procedures, and speed up production (Ballard 1999). This concept has been applied in many industries, generating a wide range of benefits, including production cost reductions, decreased production cycle times, and inventory reductions (Kotelnikov, 2006).

Lean production principles have been successfully applied to the construction industry, leading to a streamlined process that reduces costs. This process requires transparency of timely information (Womack and Jones 1996). For tunnel construction, given the large uncertainties of ground conditions, transparency of timely information is especially important and involves the frequent updating and delivery of modified schedules and multiple metrics to multiple stakeholders.

Decision Aids for Tunneling (DAT)

The Decision Aids for Tunneling (DAT) allow project managers to simulate tunnel construction, taking into account uncertainties in ground conditions for a given tunnel; as a result, managers obtain distributions of the total cost and duration of tunnel construction (Sinfield and Einstein 1996). To consider the uncertainties, geologists and engineers produce probabilistic geologic/geotechnical profiles that indicate the probabilities of particular geologic conditions occurring at a particular tunnel location (Einstein 2001). The profiles for all parameters are then combined in ground class profiles, such as RMR and Q systems. A number of such profiles are simulated to represent the whole range of geologic conditions.

Observations in the excavated area often allow the project managers to refine or update predictions for the unexcavated part (Haas and Einstein 2002). As excavation progresses, predictions about the remaining part of the tunnel must be periodically updated to allocate resources and reduce waste. As a result, the updating module of the DAT was developed, relying on a Bayesian updating procedure and direct updating information. However, the DAT does not enable project managers to automatically update ground conditions and the consequent quantities of the support system for the unexcavated sections of tunnel. In addition, the DAT's static analysis process cannot support the managers in making timely and consistent decisions during the construction phase.

Virtual Design and Construction (VDC)

VDC is the use of integrated multi-disciplinary performance models of design-construction projects to support explicit and public business objectives (Kunz and Fischer 2009). Users can implement three distinct phases to implement the VDC—visualization and metrics, integration, and automation—but this study focuses on the first phase to investigate the feasibility of the VDC as a decision-support system for tunnel construction.

4D model-based analysis tools are appropriate VDC tools for implementing the first phase of the VDC. A 4D model, which links the objects of the 3D CAD model with the activities of the time schedule, is used for the visualization (Bazjanac 2004); the 4D model is also useful for a wide range of other purposes, including site layout plans and cost estimations (Hartmann et al. 2008). Therefore, the 4D model will be able to support better decision-making for tunnel construction schedules by analyzing and representing the schedule and consequent resource plans simultaneously. In the following section, a 4D model-based decision-support system, the Development Strategy Simulator (DSS), is applied to a U.S. tunnel construction project to evaluate the usefulness of the DSS for tunnel construction.

Application of the DSS

This section introduces the DSS, which visualizes construction schedules and analyzes multiple performance metrics simultaneously. The main purpose of the DSS in this paper is to provide timely and accurate information to determine the schedule, resulting in resource reallocation. To this end, the following three objectives are established:

(1) Visualization of tunnel construction schedule without a manual linking process
(2) Analysis of multiple metrics simultaneously
(3) Representation of multiple metrics at various points in time
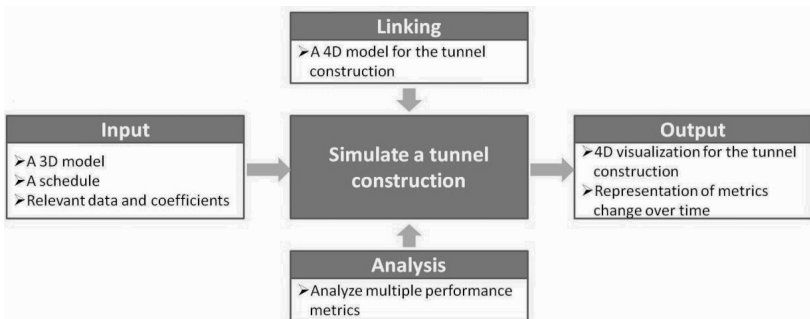
Figure 3. The analysis process of the DSS for tunnel construction

The DSS analysis consists of four steps. First, users create a 3D model and a schedule. Next, they define relevant data and coefficients for performance metrics analysis. Then, they create a 4D model, and the DSS analyzes the metrics required. Finally, the DSS visualizes the tunnel construction plan and represents multiple metrics (see Figure 3).
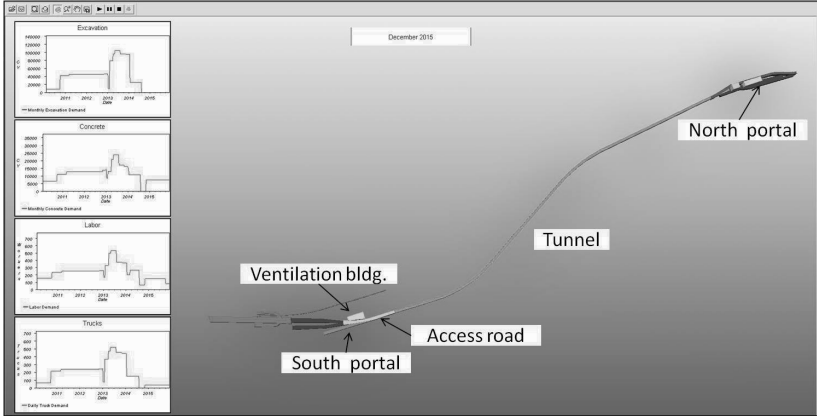
Figure 4. A snapshot of the DSS applied in this study

In this study, the DSS was applied to a U.S. road tunnel construction project in Seattle, WA. The project included four elements: tunnel, portals, access road, and ventilation building. The project manager provided 2D drawings, milestones for each element, and quantities required for each element. Based on this information, a 3D CAD model and schedule were developed. In addition, the DSS allowed for the direct input of the quantity information for each element. In this application, one specific set of ground conditions was considered to evaluate the DSS's capability of timely and transparent information delivery. The consistent naming convention in the 3D model and activities enabled the objects of the 3D CAD model to be linked with the activities of the schedule without requiring a manual process. Following the analysis process of the DSS, the tunnel construction was simulated.

The DSS visualizes the construction schedule and represents multiple metrics simultaneously (see Figure 4). Visualization of the tunnel construction enables more stakeholders to participate in the project review. Metrics (e.g., the number of trucks, amount of concrete, hours of labor, and quantity of excavation) analyzed by the DSS speed up decisions to allocate resources with other stakeholders. In addition, if the coefficients can be defined, the DSS should be able to analyze whatever metrics the users want to evaluate.

Discussion

The DSS supports multiple stakeholders in making timely decisions on their tunnel construction schedules by providing them with visualization of the construction schedule and simultaneous representation of multiple metrics required

for the schedule. Consequently, the DSS supports quick schedule modifications and resource reallocation, taking into account more possibilities from updated predictions of ground conditions throughout the entire construction period. In addition, the DSS supports multiple stakeholders in undertaking a tunnel project based on lean construction principles by providing multiple metrics simultaneously. As a result, the DSS supports multiple stakeholders in carrying out their project in an efficient and effective way in terms of schedule and cost for a tunnel project undertaken with uncertain ground conditions.

However, the DSS analysis showed that the third phase of the VDC (i.e., automation) should be implemented for making more timely and robust decisions on tunnel construction schedules. To this end, there are two remain challenges. First, project managers have difficulty with automatically and consistently predicting multiple possible sets of ground conditions at different points in time. Also, the managers have difficulty generating and evaluating multiple alternative construction schedules in a timely and consistent manner.

To overcome such challenges, the following things are still required: (a) a formalized mechanism to generate a progress-adaptive stochastic earth model for multiple possible sets of ground conditions; (b) a formal representation of multiple construction methods, taking into account construction schedule changes; and (c) a formalized mechanism to quickly generate and evaluate multiple alternative schedules.

References

Ballard, G. (1999). "Improving work flow reliability." *Proc. 7th Int. Group for Lean Construction Conf.*, Univ. of California at Berkeley, Berkeley, Calif., 275–286.

Barton, N., (1995)." The influence of joint properties in modeling jointed rock masses." Keynote Lecture. In: Proc. 8th Congress of ISRM, vol. 3, Rotterdam Balkema.

Bazjanac, V. (2004). "Building energy performance simulation as part of interoperable software environments." *Building and Environment*, 39, pp. 879-883.

Bieniawaski, Z.T., (1989). "Engineering Rock Mass Classifications. A complete manual for engineers and geologists in mining, civil and petroleum engineering." John Wiley and Sons, Inc., p. 251.

Einstein, H.H. (2001). "The decision aids for tunnelling (DAT) – A brief review." *Tunnelling Technology*, Korea Tunnelling Assoc., Vol. 3, pp. 37-49

Hartmann, T., Gao J. and Fischer, M. (2008). "Areas of application for 3D and 4D Models on construction Projects." *Journal of Construction Engineering and Management*, 134(10), pp. 776-785.

Haas, C. and Einstein, H.H. (2002). "Updating the decision aids for tunneling." *Journal of Construction Engineering and Management*, ASCE, Vol. 128, pp. 40-48.

Kotelnikov, V. (2006). "Lean production: Doing more with less." (http://www.1000ventures.com/business_guide/lean_production_main.html)

Kunz, J. and Fischer, M. (2009). "Virtual design and construction: themes, case studies and implementation suggestions." *CIFE Working Paper #097*, Center For Integrated Facility Engineering, Stanford University, USA, 2005.

Panthi, K. K., and Nilsen, B. (2007). " Predicted versus actual rock mass conditions: A review of four tunnel projects in Nepal Himalaya." *Tunnelling and Underground Space Technology 22,* 173-184.

Sinfield, J.V. and Einstein, H.H. (1996). "Evaluation of tunneling technology using the decision aids for tunneling." *Tunnelling and Underground Space Technology*, Vol. 11, pp. 491-504.

Womack, J. P., and Jones, D. (1996). "Lean thinking: Banish waste and create wealth in your corporation." Simon & Schuster, New York, 352.

Porosity Distribution and Flow Rate of Granular Materials

J. A. Wasserman[1], Lorin Nickle[2], Ali Daouadji[3], and Beena Sukumaran[4]

[124]Rowan University, Department of Civil and Environmental Engineering, 201 Mullica Hill Road, Glassboro, NJ 08028, USA; PH (856)256-5324; FAX: (856)256-5242; e-mail: sukumaran@rowan.edu

[3]Laboratoire d'Etude des Microstructures et de Mécanique des Matériaux LEM3, UMR CNRS 7239, Université Paul Verlaine – Metz, Ile du Saulcy, 57045 Metz, France

ABSTRACT

The flow rate of granular material depends on particle characteristics including mineralogy, shape, angularity, surface texture, grain size and size distribution. This paper discusses computational and numerical techniques used to measure material characteristics (input) as well as computational experimentation and measurement techniques used to evaluate material flow rate (output) from a hopper. X-Ray Computational Tomography (XCT) image analysis is used to measure sample porosity and porosity distribution in both 2D and 3D, while microscopic image analysis methods are used to compute the angularity and shape factor of sample particles. With these input characteristics, we can evaluate the flow rate of a sample using Discrete Element Method (DEM) Modeling, which can then be compared to and verified by physical experimentation.

INTRODUCTION

The behavior of granular material in hopper flow is affected by many inherent particle characteristics including mineralogy, shape, angularity, surface texture, grain size and size distribution. Standard methods exist to quantify grain size and size distribution (ASTM D422 and D854), but such methods have not been developed to easily quantify morphology. (Sukumaran and Ashmawy 2003) discuss a combination of imaging and analytical analysis techniques used to quantify the morphology of a particle. (Sukumaran 1996) outlines a flow rate experiment used to develop a correlation between the various output parameters (mass flow rate, volume flow rate and solid volume flow rate) and inherent particle characteristics (angularity and shape). Experimental flow rate results from Sukumaran are discussed and will be used as verification for 3-D DEM. (Nedderman 1981) reviews articles discussing correlations to approximate flow rate based on input variables including sample density and orifice diameter, as well as coefficients determined for different sand types. In (Bloom, Russell, and Kustau 2010), various image processing techniques are combined with XCT imaging to generate 2D and 3D plots of sample porosity. (To and Lai 2002) uses numerical analysis to investigate 2D granular flow,

specifically jamming factor related to the ratio between particle diameter and orifice diameter. This experimental data is used to verify 2-D DEM results.

To model the flow of granular particles, Particle Flow Code (PFC) 2D (Itasca 2008a) and 3D (Itasca 2008b) are used to model the flow through a hopper. The environment is controlled by governing physics calculated at an interval time step for each particle contact. Since the software calculates values for every particle, it is necessary to provide accurate information about the sample. Experimentation in PFC can be used to more quickly validate correlations found in (Nedderman 1981) than performing time consuming physical experiments. It has been shown based on earlier studies (Sukumaran 1996 and Sukumaran et al. 2006) that flow rate through a hopper can be used as a measure of the shear strength properties of granular material. This study will verify this using both experimental data and numerical modeling results.

METHODS

Various imaging methods are used to determine key sample properties that otherwise may be more difficult to calculate, like particle morphology and porosity. Computational modeling techniques are used to recreate flow rate experiments, allowing for precise control over all parameters of a sample. This form of experimentation allows us to conduct flow simulations more frequently than physical experimentation, while the parameter control aspect allows us to more easily correlate a change in flow rate to a specific input parameter.

IMAGE ANALYSIS AND MORPHOLOGY (2D)

To quantify the morphology of a particle, a 2D projection of the particle is analyzed for its sharp edges and general shape. For each particle, an Angularity Factor is calculated on the number and sharpness of corners, and a Shape Factor is defined as the deviation of the particle outline from a circle.
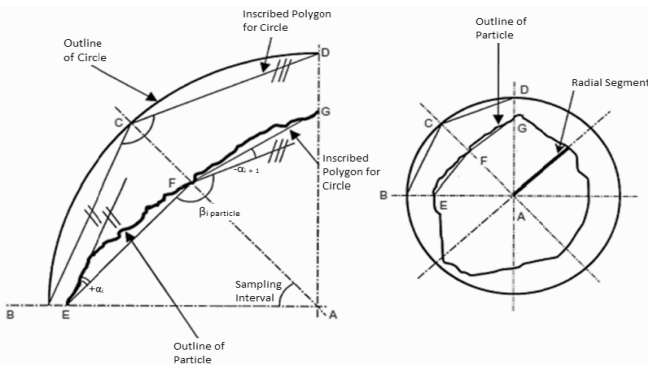
Figure 1: Geometric shape used to verify and calculate shape and angularity factor.

To quantify particle morphology, the normalized shape factor is calculated from Equation 1, and the angularity of a particle is calculated from Equation 2. Particle Shape Factor is described by the distortion angle α_i, which denotes the difference between the angle of a particle and that of a circle as shown in Figure 1. Particle Angularity is described in terms of the number and sharpness of the corners of the particle. The sum of the difference between these β angles, shown in Figure 1, determines the particle angularity (Sukumaran and Ashmawy 2003). This equation gives AF = 0% for a spherical particle and AF = 100% for the sharpest particle analyzed.

Equation 1: Normalized Shape Factor

$$SF = \frac{\sum_{i=1}^{N}\left|\alpha_{i\ particle}\right|}{N \times 45°} \times 100\%$$

Equation 2: Normalized Angularity Factor

$$AF = \frac{\sum_{i=1}^{N}(\beta_{i\ particle} - 180)^2 - (\frac{360^2}{N})}{3 \times (180)^2 - (\frac{360^2}{N})}$$

XCT IMAGING AND POROSITY

In addition to conventional experimental methods, the use of an X-Ray Computational Tomography (XCT) image analyzer is used to measure sample porosity and porosity distribution (Razavi, Muhunthan, and Al Hattamleh 2007). The porosity of a sample, n can be defined as the ratio of voids to total volume, shown in Equation 3 where V_v is the volume of empty space and V is the total volume of the enclosed sample.

Equation 3: Porosity, n

$$n = \frac{V_v}{V}, \qquad 0 \leq n \leq 1$$

XCT scans produce vertical slices of a 3D granular sample taken as the sample rotates around its longitudinal axis. These slices are then reconstructed into a solid model. This model can be further analyzed either as a series of images taken from horizontal slices of the model, shown in Figure 2 or a solid 3D model, Figure 3. For 2D porosity, a MatLab program crops an area of a slice and analyzes the image by its grayscale pixels. A threshold is calculated according to (Bloom, Russell, and Kustau 2010) to convert the grayscale image to a binary image. This threshold value is used to determine above what grayscale value from 0 to 1 represents a particle, and below which represents empty space. Thresholding is used to decrease the influence of artifacts and image blemishes. The binary pixels represent voids (black pixels, 0's) and granular material (white pixels, 1's), which are used to calculate the porosity of the sample. (Chin et al. 2012)
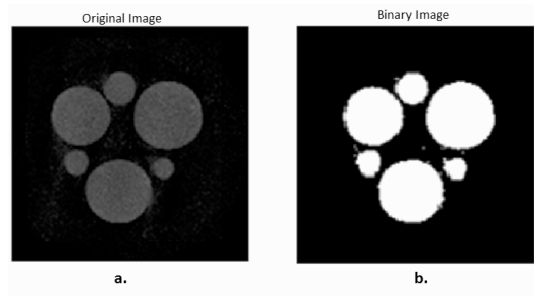
Figure 2: Reconstructed XCT images of 4mm glass beads.
a. Original Image b. Binary version of image

For 3D porosity, the 3D assembly is imported into MatLab where a program analyzes the boundaries of the particles and is able to directly calculate the volumes.
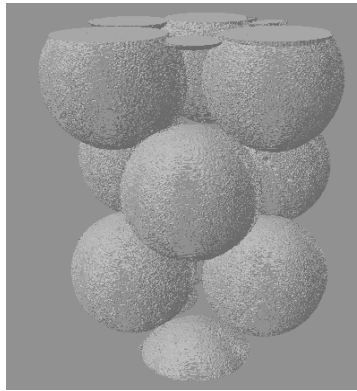
Figure 3: Reconstructed 3D model of 4mm glass beads.

PFC MODELING AND FLOW RATE

PFC modeling uses a physics controlled experimental environment to replicate a granular assembly. Balls and Discs (3D and 2D) are used as granular particles to predict the movement of individual particles or clumps. A force-contact model is applied to the assembly to calculate the forces acting at each contact on a particle. By calculating the forces at each contact for each time step, the software can predict the movement of the assembly.

To model a flow assembly, a distribution of disks is generated according to a set of input parameters; porosity, particle size distribution, min and max diameter, particle stiffness and friction coefficient. The disks are generated at random locations in a predefined generation area, however this can cause particle overlap which leads to an inaccurate porosity and an unbalanced static force. To remedy this, the disks

are generated at a smaller scale, then allowed to "grow" and fill in the area of generation. The growth rate is controlled by the ratio of generated porosity to desired porosity; particles are allowed to grow until the desired porosity is reached.
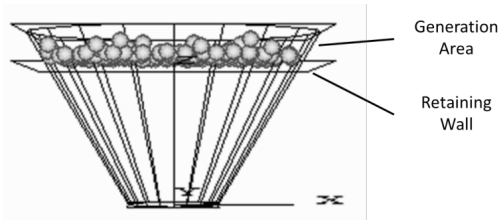
Figure 4: PFC 3D initial particle generation, 1mm glass sphere.

Initially, the assembly settles for a prescribed number of steps, where the number of steps depends on the quantity of particles generated. The number of steps required is overestimated to ensure sufficient particle settling and can later be modified to make the model more efficient. The generation area is the top portion of the flow cone with a temporary wall as the bottom boundary of the assembly shown in Figure 4. When this wall is removed, the sample drops under dry pluviation into the hopper where it will be held until a temporary wall capping the base of the hopper is removed, shown in Figure 5. This step is maintained until stresses are resolved, when the hopper cap is removed and the assembly is allowed to begin flowing. The flow rate measurement begins when the hopper orifice is opened. The measurement stops when the last particle breaks the plane of the flow cap. Since each particle can be different from one another, the program calculates the stop condition and flow conditions by cycling through each particle individually. This allows for the calculation of exact mass and solid volume of the sample, shown in Equation 4. Combining these values with the time from flow start to finish determines the Mass and Solid Volume flow rates (MFR and SVFR). Volumetric Flow Rate (VFR) can be determined by combining SVFR with measured porosity.

Equation 4: Solid Volume Flow Rate.

$$SVFR = \frac{V_{solid}}{t} = \frac{W_s}{G_s \gamma_w t} = \frac{MFR}{G_s \gamma_w} = \frac{VFR}{(1+e)}$$

Where:

W_s = weight of solids
G_s = specific gravity of solids
γ_w = unit weight of water
t = time over which flow rate is measured
e = void ratio
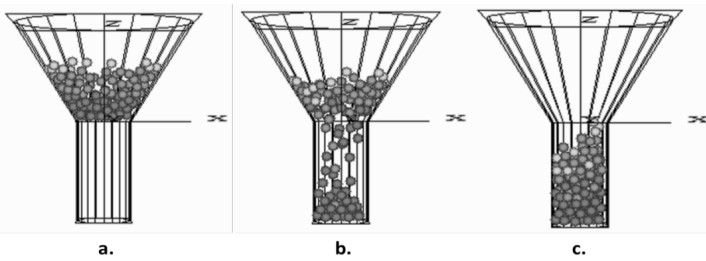
Figure 5: Granular assembly in hopper flow; a. pluviated sample, b. mid flow, c. completed flow.

RESULTS AND CONCLUSIONS

Table 1 includes morphology results according to (Sukumaran and Ashmawy 2003). Results show that spherical Glass Ballotini has the lowest angularity and shape factor of 2% and 4% respectively. More angular sands have angularity values <20%, (max 29% for Ottawa #90) and shape factor values between 30 and 50% (max 51% for Ottawa #90). These morphology results *directly* correlate to flow results collected by Sukumaran.

Table 1: Morphology analysis results for various materials.

Material Description	Shape Factor (%)			Angularity Factor (%)		
	Mean	Standard Deviation	Coefficient of Variation	Mean	Standard Deviation	Coefficient of Variation
Ottawa #45 (angular)	47	8	0.170	7	2	0.286
Ottawa #90 (angular)	51	9	0.176	29	8	0.276
Glass Ballotini #140	4	1	0.250	2	0	0.000
Daytona Beach Sand	42	9	0.214	19	7	0.368
Michigan Dune	32	6	0.190	9	2	0.250

Porosity measurements in 2D and 3D are calculated using various image-processing techniques. These techniques yield the results in Table 2 while results from methods outlined in Bloom are shown in Table 3. Results show that if an appropriate threshold is chosen for the 2D analysis, the calculated porosity will be relatively accurate, within 6% compared to the 3D porosity. The 3D results are accurate in that they are an exact analysis of the particles, however a 3D calculation cannot be done on a full sample, due to computational resources. For this reason, it would be ideal to calculate porosity in 2D rather than 3D, however more work is needed to ensure that an accurate threshold value can be calculated for each image. A

strict testing protocol for X-ray scans control the quality of images for analysis in an attempt to increase the accuracy of 2D measurements. Table 3 results also show that as the angularity and shape factor of a material increases, the porosity of the sample decreases. This is believed to be due to the angles of particles fitting together more closely than the broad angles in low morphology particles, specifically glass ballotini. Daytona Beach sand appears to be an exception to this correlation, which is believed to be due to smaller mean particle size compared to the other sands tested. The smaller particle size and particle size distribution are believed to create a denser packed sample, giving a lower porosity. To further investigate this correlation, the porosity of more material types will be sampled and analyzed for comparison.

Table 2: 2D and 3D porosity results.

Granular Material Type	Average 2D	3D	%diff
4mm Glass Beads	46.06%	47.60%	3.34%
Dry #1	39.58%	37.18%	-6.06%

Table 3: Measured porosity data from (Bloom, Russell, and Kustau 2010).

Granular Media	Min	Max	Mean	Standard Deviation
Glass Beads	0.2626	0.3494	0.3004	0.0132
Michigan Dune	0.3146	0.4177	0.3550	0.0176
Daytona Beach	0.1546	0.6173	0.2801	0.0383
Dry #1	0.2964	0.4191	0.3634	0.0201

Flow rate results according to (Sukumaran 1996) are shown in Figure 6. The results show that Glass Ballotini has the highest flow rate of the materials tested, while Ottawa #45 has the lowest flow rate. As shown in Table 1, Glass Ballotini has the lowest angularity and shape factor of samples, while Ottawa #45 is the most angular particle, showing a positive correlation between particle morphology and solid volume flow rate.

FUTURE WORK

An analysis of thresholding techniques outlined by (Bloom, Russell, and Kustau 2010) will aid in verifying the accuracy of 2D Porosity measurement techniques. Correlations between morphology and shape factor will be expanded with the testing of other granular materials. PFC modeling will be used to analyze the effects of angularity on flow rate using physical experimentation to validate the results.
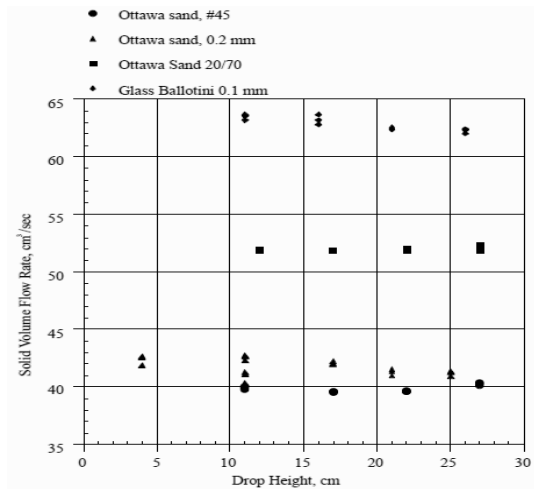
Figure 6: Solid volume flow rate vs. Drop height.

ACKNOWLEDGMENTS

REFERENCES

Bloom, Michael, Michael Russell, and Aliaksei Kustau. 2010. "Measurement of Porosity in Granular Particle Distributions using Adaptive Thresholding."

Chin, Jonathan, Nicholas Cincotti, Peter D'Amico, and Lorin Nickle. 2012. "Calculating 2D and 3D Porosity and Using Discrete Element Methods to Identify Shear Strength and Granular Flow through a Hopper."

Itasca. 2008a. "PFC 2D 4.00-123."

———. 2008b. "PFC 3D 4.00-121."

Nedderman, R. M. 1981. "The Flow of Granular Materials." 1597 – 1609.

Razavi, Mohammad Reza, Balasingam Muhunthan, and Omar Al Hattamleh. 2007. "Representative Elementary Volume Analysis of Sands using X-Ray Computer Tomography." 212 – 219.

Sukumaran, Beena. 1996. "Study of the effect of particle characteristics on the flow behavior and strength properties of particulate materials."

Sukumaran, Beena, and Alaa K Ashmawy. 2003. "Influence of inherent particle characteristics on hopper flow rate." Powder Technology 138(1). http://linkinghub.elsevier.com/retrieve/pii/S0032591003002304.

To, Kiwing, and Pik-Yin Lai. 2002. "Jamming pattern in a two-dimensional hopper." Physical Review E 66(1): 1–8. http://link.aps.org/doi/10.1103/PhysRevE.66.011308 (October 4, 2012).

From 3D tomography to physics-based mechanics of geomaterials

I. Vlahinić[1], J.E. Andrade[1], E. Andó[2], G. Viggiani[2]

[1]California Institute of Technology, Department of Mechanical and Civil Engineering,
Pasadena, CA 91125, USA
[2]Université Joseph Fourier, Laboratoire Sols, Solides, Structures – Risques,
38041 Grenoble, France

ABSTRACT

Computed tomography (e.g. X-ray CT) has provided unprecedented 3D images of microstructures, with the scale of individual voxels on the order of a few micrometers. The technique has been especially powerful in granular media, where a time sequence of snapshots containing thousands of grains can be collected in situ, for a given load path. In turn, this has opened up great opportunities in relating the changes in granular microstructure to average mechanical properties and has offered a clear path to move beyond phenomenological constitutive models.

A shortcoming of X-ray CT is that it provides images or photographs, which then must be converted into usable geometric data. Here, we introduce a framework for this purpose, which we term 'Level Set Bridge'. The proposed level set framework provides a basis for a comprehensive geomaterial characterization, and ultimately serves a direct link to computational methods in geomechanics. We briefly showcase the technique through examples of a natural material, Caicos ooids, consisting of highly rounded particles. In particular, we use X-ray CT images of ooids under a triaxial load sequence to quantify local grain-scale geometry and kinematics, and chart a path toward extraction of granular fabric.

MOTIVATION

Classical geomechanical laboratory approaches, e.g. a triaxial apparatus, consider a relatively uniform material loading (usually with a confining pressure), and measure forces and displacements at the boundary. This has historically provided an excellent quantitative indication of the bulk material response. And no one has described the observed bulk response of geomaterials better than professor emeritus of civil and mechanical engineering at Caltech, Dr. Paul Jennings, who at a memorial gathering of another great, Dr. Ronald Scott, characterized geomaterials as "noneverything: nonlinear, nonelastic, nonhomogeneous, nonisotropic, and, therefore, noneasy." This description arises because geomaterials invariably display a great deal of grain-scale complexity, which is where their macroscopic behavior is encoded. For example,

fractured rocks usually become stiffer and their elastic modulus increases under increasing confining pressures, a physical result of closing of the micro cracks (Nur and Byerlee, 1971). Another canonical example is that of sand, in which re-arrangement of grains due to imposed loads is often non-recoverable, and thus overwhelmingly plastic. The latter serves as an example geomaterial in this work.

At material scales of the order of sand particles, approximately 100 micrometers and above, gravity and (if fluids are present) capillary forces dominate (Santamarina, 2003). It is at these scales that X-ray CT experiments continue to provide unprecedented insight into the material behavior of granular materials. It is also in this regime that grain topology has a great impact on the overall constitutive behavior and strength. This confluence, namely the ability to truly see the material scale of importance and interest in materials such as sand, furnishes the motivation for this work.

GRAIN-SCALE GEOMETRIC PROPERTIES

In a contribution currently under review (Vlahinic et al., 2012), we introduced a level set methodology to extract individual microstructural features from X-ray CT images. Specifically, we focused on characterization of individual grains using a level set methodology. Discussed were quantitative complexities and solutions related to discriminating grains from voids, and also grains from other neighboring grains in images, i.e. identifying grains as individual units. Figure 1 provides a visual schematic of key non-trivial steps involved in the process, with a sampling of the results on a natural geomaterial, the Caicos ooids. Ooids present an attractive model material, owing to the highly rounded nature of their individual grains.

Here we use as a starting point a level set of a grain, and extend our previous work to examine physical properties of grains. First, we show how the basic properties of the level set (LS) (Osher & Fedkiw, 2002) allow us to easily calculate grain surface properties, and also locate the points of grain-grain contact. The latter in particular, is crucial for extracting the material fabric, a statistical measure of relative contact densities and their orientations.

The advantage of LS in describing complex geometric shapes is that it resides on a regular Cartesian grid *identical* to that of the underlying image. In this way, different grain shapes simply take on different functional values, while the underlying computational grid remains the same. Grain surface Γ is conveniently defined as a zero-level contour of the level set function $\varphi(\boldsymbol{x})$, where $\boldsymbol{x}$ describes a 3D domain. In the grain interior, namely in the region Ω^-, $\varphi(\boldsymbol{x})$ is negative, while in the grain exterior, namely in the region Ω^+, $\varphi(\boldsymbol{x})$ is positive, such that

$$\varphi(x) < 0, \text{ region inside grain surface } \Gamma, \text{ denoted by } \Omega^-$$
$$\varphi(x) = 0, \text{ grain surface } \Gamma, \text{ denoted by } \partial\Omega \qquad (1)$$
$$\varphi(x) > 0, \text{ region outside grain surface } \Gamma, \text{ denoted by } \Omega^+$$

In principle, any function satisfying the above criteria can be used as an LS function. However, a signed distance function is typically adopted because it provides the following convenient properties:

$$|\nabla\varphi| = 1; \; \boldsymbol{n} = \nabla\varphi/|\nabla\varphi| = \nabla\varphi; \; \kappa = \nabla \cdot \boldsymbol{n} = \Delta\varphi \tag{2}$$

where ∇ and Δ are the gradient and laplace operators, while $\boldsymbol{n}$ and κ are the vector normal and curvature. Evaluated on the surface of a grain, at locations where $\varphi(\boldsymbol{x}) = 0$, quantities $\boldsymbol{n}$ and κ provide a detailed topological grain surface map. Additionally, if locations of grain-grain contacts can be found, Eq. 2 provides an excellent platform for quantifying grain fabric.

Indeed, determining grain-grain proximity, and thus finding grain-grain contact locations, also turns out to be relatively simple using the LS methodology. Let's assume that two LS function are available, with φ_1 associated with the first grain and φ_2 associated with the second grain. In this case, the location on a Cartesian grid where two the grains are closest to each other, a numerical indicator of grain-grain contact, is simply the minimum of the function f($\boldsymbol{x}$), where:

$$\mathbf{f}(\boldsymbol{x}) = \; \max\big(\boldsymbol{\varphi_1}(x), \boldsymbol{\varphi_2}(x)\big) \tag{3}$$

While we have not yet assembled a full statistical description of contact properties of the representative material volumes for the granular ooids discussed in this work, we are well on our way. The results of this analysis should be ready to present at the 2013 ASCE International Workshop on Computing in Civil Engineering, the venue for which this conference paper is intended. If successful, the results would present a transformative advance over previous experimental characterizations of grain fabric of natural geomaterials, first pioneered by Oda and colleagues (e.g. Oda, 1972), using a post-mortem analysis of triaxial specimen at various stages of deformation.
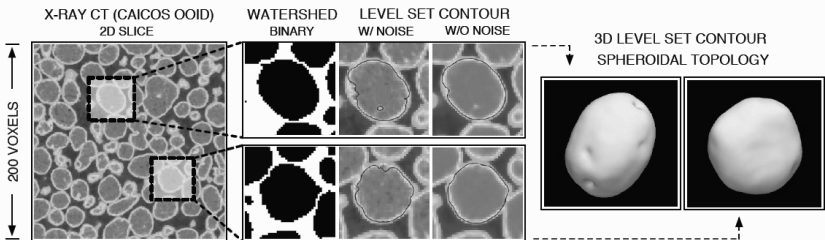
Figure 1. Level set based characterization of a natural geomaterial - highly rounded Caicos ooids – via 3D X-Ray CT experiments. [Left to right] A series of steps leading from a raw image volume (slice shown) to a full 3D rendering of sand-grain surfaces (displayed without *any* topological modifications) using the Level Set methodology. Adopted from Vlahinić et al. (2012).

GRAIN-SCALE KINEMATICS

Here we discuss how the LS-based framework can be used to examine a time series of X-ray images in order to determine grain-scale kinematics. An example of the results is shown in Figure 2 where the same grain is extracted from X-ray CT images at two different load stations – from a region that will ultimately belong to the interior of a localization band. This region is characterized by substantial grain kinematics, in terms of both translational and rotational motions.

As a starting point for determining grain kinematics using the LS framework, we begin by determining the grain centroid x_c at each image station:

$$x_c = \left. \sum H\big(-\varphi(x)\big)\, x\, dx \middle/ \sum H\big(-\varphi(x)\big)\, dx \right. \tag{4}$$

where $H(-\varphi)$ represent a Heaviside of negative of the LS function[1]. The distance between the grain centroids at different image stations defines the grain translation, such that $t = x_{c2} - x_{c1}$. Furthermore, a rotation matrix R is sought such that all vectors $m^{\alpha 1}$ emanating from the centroid to all points in the volume region Ω^- at an initial load station can be mapped onto the corresponding vectors $m^{\alpha 2}$ at any other load station. Formally, this relation is expressed for all α vectors, such that $m^{\alpha 2} = t + R \cdot m^{\alpha 1}$. In reality, all vectors $m^{\alpha 1}$ will not map exactly onto all vectors $m^{\alpha 2}$. Instead a cost function C in Eq. 5 needs to be defined, such that a minimum of C yields the most likely rotation vector R that maps a grain at one time station onto the same grain at another time station (Smith et al., 2002).

$$C(R) = \tfrac{1}{2}\sum_\alpha [\varphi(m^{\alpha 2}) - \varphi(m^{\alpha 1})]^2 \tag{5}$$

To determine the solution to Eq. 5, it helps to define the rotation matrix using an axis-angle representation, via a simplified Rodrigues rotation formula, such that in indicial notation,

$$R(r) = R_{ij}(r) = (1 - r_l r_l)\,\delta_{ij} + 2 r_i r_j - 2\sqrt{1 - r_l r_l}\,\epsilon_{ijk} r_k \tag{6}$$

where $r = r_i = \sin\left(\tfrac{\theta}{2}\right) b$, θ is the magnitude of the rotation, and b is the unit vector parallel to the axis of rotation; δ is Kronecker delta and ϵ is the permutation (Levi-Civita) tensor. The cost function can then be minimized in terms of the three Cartesian components of the unknown vector r (rather than the matrix R). The speed of minimization can be improved by analytical evaluation of gradient of (5).

Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher, 1987), a quasi-Newton method, requires on average 5-25 iterations to reach the final solution,

---

[1] The Heaviside function is most often provided in 'smooth' form. For details, the reader is referred to any reference text on the level set method, e.g. Osher & Fedkiw, 2002.

depending on the magnitude of the rotation angle. In tests using a Matlab 2012a implementation ('fminunc') on a standard Pentium i7 laptop, even for very large 3D angular rotations of up 90 degrees, we find that the method converges to a correct solution in a fraction of a second. We attribute the fast convergence to a smooth profile of the LS function that characterizes the grains.
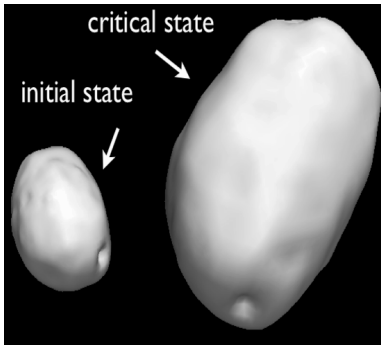
Figure 2. Grain-scale kinematics via 'Level Set Bridge'. Displayed in perspective view are two grains, independently extracted from X-ray images at two distinct load states: an initial (pre-load) and a post-localized (critical) state. The proposed platform enables the determination of grain kinematics for arbitrarily large time steps, with great accuracy. In this case, the grain translated nearly 100 voxels (approximately 1.5 mm) and rotated 71 degrees around an axis nearly normal to the plane of localization (out of shear-plane rotation).

LINKING X-RAY CT WITH CONTINUUM GEOMECHANICS

Continuum models, or more precisely continuum constitutive models, are (with a few exceptions) phenomenological in nature. They are based on empirical evidence, inferred through extensive laboratory testing or some general field observations. Theory of elasto-plasticity provides a popular and general framework for geomaterials, e.g. see Vermeer and de Borst (1984), with classic critical state models among the simplest, and thus the most instructive. What is clear, however, is that little can be learned directly in the way of physical processes at play or the causality between the processes and the parameters from continuum models alone. Overcoming this fundamental shortcoming, to yield physics-infused continuum models or to advance multi-scale approaches that link the micro and macro scales in geomaterials is the motivation and transformational potential behind the X-ray CT experiments and the proposed 'Level Set Bridge' framework, as schematically shown in Figure 3.

There is already evidence that revolution in modeling is on the way. With improvements in computational power, and widespread use of Discrete Element Methods (DEM), for example, more recent continuum descriptions of e.g. sands have started to incorporate grain-scale insight to develop much more robust, and physically consistent constitutive laws. A notable example is a model by Li and Dafalias (2010, 2012) that incorporates an anisotropic, statistical measures of pore geometry, termed 'pore fabric', a counterpart to grain fabric, to explain trends in historical experimental data with great simplicity. What is lacking, however, is quantitative insight beyond computational DEM, namely insight into fabric evolution in natural geomaterials (as a complement to numerical studies).

Ultimately, the fruits of the effort in correlating micro- and macro- fields in geomaterials show immense future promise not only in terms of material characterization. The other aspect of this work is related to verification & validation approaches for grain-scale computational models such as DEM. One DEM approach, in particular, using a non-rational b-splines (NURBS) that can incorporate truly *arbitrary* grain shapes (Andrade et al., 2012), promises to establish a close link between the grain-scale experimental and computational approaches in geomaterials.
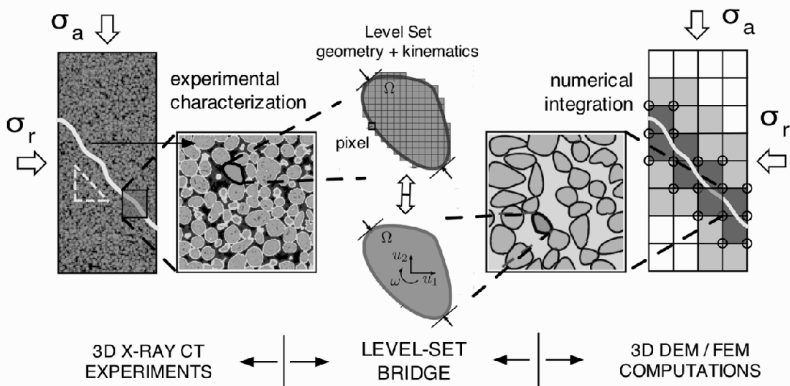
Figure 3. A schematic identifying the level set (LS) platform as a bridge between advanced experiments and modern computational approaches in geomaterials. The LS method serves as a key tool enabling accurate extraction of micro-scale geometry and kinematics from advanced X-ray experiments. Relative flexibility and simplicity of LS in describing complex geometric surfaces found in natural geomaterials stands to provide a direct link to modern modeling tools, such as the discrete element method (DEM) and the finite element method (FEM), to study micromechanical processes.

REFERENCES

Andrade, J.E., Lim, K.-W., Avila, C.F., and Vlahinić, I. (2012). Granular element method for computational particle mechanics. Computer Methods in Applied Mechanics and Engineering 241, 262-274.

Fletcher, R. Practical methods of optimization (2nd ed.). New York: John Wiley & Sons (1987).

Li, X.S. and Dafalias, Y.F. (2010). Anisotropy at critical state: The role of fabric." Proc., 9th HSTAM Int. Congress on Mechanics, The University of Cyprus, Nicosia, 57–60.

Li, X.S. & Dafalias, Y.F. (2012). Anisotropic Critical State Theory: Role of Fabric. Journal of Engineering Mechanics 138(3), 263–275.

Nur, A. & Byerlee, J.D. (1971). An exact effective stress law for elastic deformation of rock with fluids. Journal of Geophysical Research 76 (26) 6413-6419.

Oda, M. Initial fabrics and their relations to mechanical properties of granular material. Soils and Foundations 12, pp. 17-36 (1972).

Osher, S.J. and Fedkiw, R.P. Level Set Methods and Dynamic Implicit Surfaces. Springer, 1st edition (2002).

Santamarina, J.C. Soil behavior at the microscale: Particle forces. In Proc. ASCE, Soil Behavior and Soft Ground Construction, pages 25-56, 2003.

Smith, T.S., Bay, B.K., and Rashid, M.M. Digital volume correlation including rotational degrees of freedom during minimization. Experimental Mechanics 42, pp. 272-278 (2002).

Vermeer, P.A. and de Borst, R. Non-associated plasticity for soils, concrete and rock. Heron 29, pp. 1–62 (1984).

Vlahinić, I., Andó, E., Viggiani, G. and Andrade, J.E. Towards a more accurate characterization of granular media: extracting quantitative descriptors from grain-scale images. In review (submitted, Sept. 2012).

A time-domain substructuring method for dynamic soil-structure-interaction analysis of arbitrarily shaped foundation systems on heterogeneous media

Chanseok Jeong[1], Elnaz Esmaeilzadeh Seylabi[2], and Ertugrul Taciroglu[3]

Civil & Environmental Engineering Department,
University of California, Los Angeles, Los Angeles, CA 90095-1593

ABSTRACT

Soil-structure-interaction (SSI) effects are typically non-negligible for structures that possess one or more of the following attributes: a massive super-structure and/or foundation elements, a large uninterrupted footprint or interface with the soil domain, a soft supporting soil medium. The SSI effects comprise the dynamic interaction between the far-field soil domain, the—potentially inelastically behaving—near-field soil domain, and the structure. The far-field domain is semi-infinite unless the bedrock, or a rock outcrop is very near, and thus it can be represented with a reduced-order model in the form of impedance functions. The use of impedance functions in SSI analyses allows the computational cost to be reduced by several orders of magnitude, without compromising the solution accuracy. Moreover, it is now possible to obtain time-domain representations of the inherently frequency-dependent impedance functions. As such, accurate nonlinear time-history analyses of problems that involve SSI effects can be now carried out in a computationally efficient manner. However, the current catalogue of impedance functions is limited to simple foundation shapes and soil profiles. In the present study, we provide a systematic approach with which impedance functions for arbitrarily shaped foundations resting on (or embedded in) heterogeneous soil domains can be obtained. In order to obtain the impedance functions, forward wave propagation analyses are carried out on a high-performance computing platform. The finite element method is employed to account for the arbitrary heterogeneity of soil, and for different foundation types and geometries. In the forward analyses, the semi-infinite remote boundaries are treated with Perfectly Matched Layers (PML), which, to date, are considered to offer the best Wave-Absorbing Boundary Condition (WABC) representation. Practical examples are provided that display pronounced variations in impedance functions with respect to frequency, which illustrate and quantify the importance of using frequency-dependent impedance functions in SSI analyses.

---

[1] Post-doctoral Research Associate (chanseok@ucla.edu)
[2] Graduate Student Researcher (elnaz@seas.ucla.edu)
[3] Associate Professor (etacir@ucla.edu)

INTRODUCTION

Analysis of dynamic soil-structure-interaction (SSI), induced by earthquake motion, is essential to the seismic design and performance assessment of many types of structures (tall buildings, bridges, tunnels, earth-retaining structures, etc.). If the SSI effects are ignored, or poorly estimated, critical response measures could be under- or over-estimated, which in turn, may compromise the safety or result in an overly conservative design.
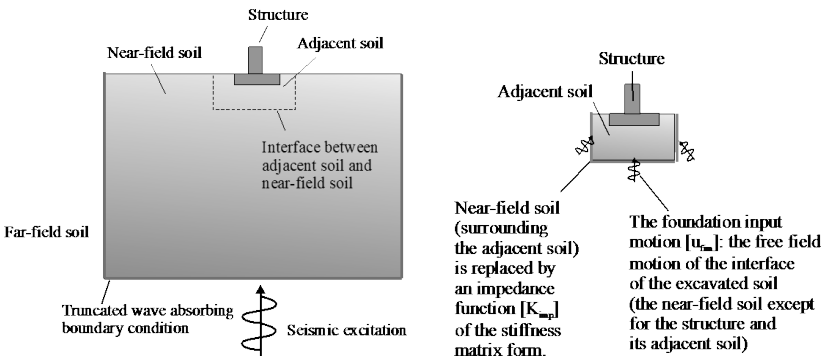
Figure 1. (a) The direct method versus (b) the substructure method.

In general, SSI analyses are conducted numerically by means of either the direct method (Fig. 1a) or the substructure method (Fig. 1b). In the direct method (Bielak, 1984; Wolf 1985), the near-field soil as well as the structure and its adjacent soil are discretized, and appropriate Wave-Absorbing-Boundary-Conditions (WABCs) are employed at the remote boundaries to truncate the semi-infinite soil domain. Seismic input motions are applied to the system by either applying accurate "jump conditions" along an arbitrary fictitious interface within the near-field soil (Bielak and Christiano, 1984) or, less accurately, by directly prescribing the separately calculated free-field displacements along truncated boundaries (Wolf, 1985). The direct method has the advantage of being directly applicable in nonlinear time domain analyses. However, its computational cost is often too high, and as such, it is not commonly attempted in engineering practice.

The substructure method (Wolf, 1985, 1989; Mylonakis et al. 1997) can be used by defining an appropriate set of impedance functions at the interface of the near-field soil and the adjacent soil (or the foundation). Seismically induced input motions are prescribed along the interface of the near-field domain and the adjacent soil, or directly at the foundation. These motions are typically computed using "site response" analyses, which yield free-field motions, and for certain cases, they are modified to account for the so-called "kinematic" effects (see, for example, Gazetas, 1984)—i.e., the effects of the stiffness and geometry of the foundation on the incoming seismic waves. The substructure method is typically carried out via

frequency-domain analyses, which have low computational cost, but are nominally restricted to the linear response regime. The method can also be adopted for the nonlinear time-history analysis in the time-domain, provided that the frequency-dependent impedance functions can be transformed to the time domain—see, for example, the rational approximation technique proposed by Safak (2006).

To date, impedance functions that are available in open literature are typically for simple soil profiles and rigid foundations. These functions are obtained via analytic or semi-analytic methods. In order to utilize the substructure method for a general SSI problem, it is necessary, among other things, to have the tools to obtain impedance functions for arbitrary soil profiles and foundations of different types (rigid, flexible, surface, embedded, void, etc.). In this paper, we present a procedure for numerically computing the impedance functions for a more general setting wherein *the foundation is arbitrarily shaped,* but *rigid*, and *the soil profile domain is arbitrarily heterogeneous*.

We employ the finite element method (FEM) to discretize the near-field soil. The semi-infinite far-field is represented using proper Perfectly Matched Layers (PML), which, to date, are considered to provide the best WABC regardless of the incident angles and frequencies of outgoing propagating waves (Kucukcoban and Kallivokas, 2011). A time-domain wave solver is developed to compute the reaction forces for a given displacement time history. Then, the impedance matrix is computed for the frequency spectrum included in the loading time history. The method is verified against known solutions, and the effects of material heterogeneity on the impedance functions are explored.

A NUMERICAL APPROACH TO DETERMINING IMPEDANCE FUNCTIONS
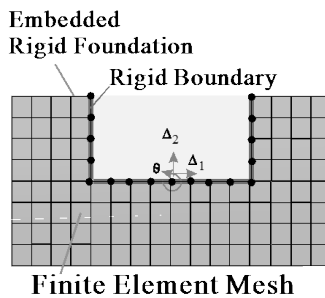
Figure 2. A rigid embedded foundation in a 2D setting (the grid indicates the discretized finite element meshes and the dots indicate the corresponding nodes on the excavated foundation boundary).

For a rigid foundation in a 2D setting (Fig. 2), the motion of the foundation can be described by three representative degrees of freedom—*viz.*, the vertical, horizontal, and rotational motions—which are denoted here as Δ_2, Δ_1, and θ, respectively. The

displacement-reaction relationship of the rigid foundation in the frequency domain can be described by using the non-dimensionalized form of an impedance matrix as in:

$$
\begin{bmatrix} \hat{R}_2 \\ \hat{R}_1 \\ \hat{M}/b \end{bmatrix} = \pi\mu \begin{bmatrix} K_V & 0 & 0 \\ 0 & K_{HH} & K_{HM} \\ 0 & K_{MI} & K_{MM} \end{bmatrix} \begin{bmatrix} \hat{\Delta}_2 \\ \hat{\Delta}_1 \\ b\hat{\theta} \end{bmatrix},
\tag{1}
$$

where "^" denotes the Fourier-transform of the subtended variable. The parameter b denotes the width of the foundation, and is used to normalize the variables $\hat{M}$ and $\hat{\theta}$, respectively, which denote the moment and the angle of rotation with respect to the foundation's centroid.

In what follows, we present a numerical procedure for computing the impedance matrix, shown in Eq. (1), for various soil profiles and different shapes of the foundation boundary. The numerical procedure entails the following steps: (1) we compute the vertical reaction $\hat{R}_2$ for a given vertical motion ($\hat{\Delta}_1 = 0$, $\hat{\Delta}_2 \neq 0$, and $\hat{\theta} = 0$) and then compute K_V; (2) we obtain the reactions $\hat{R}_1$ and $\hat{M}$ for a given horizontal motion ($\hat{\Delta}_1 \neq 0$, $\hat{\Delta}_2 = 0$, and $\hat{\theta} = 0$) and then compute K_{HH} and K_{MH}; and finally, (3) we compute the reactions $\hat{R}_1$ and $\hat{M}$ for a given rotational motion ($\hat{\Delta}_1 = 0$, $\hat{\Delta}_2 = 0$, and $\hat{\theta} \neq 0$) and then compute K_{HM} and K_{MM}.

To this end, for prescribed motions—i.e., $\Delta_1(t)$, $\Delta_2(t)$, and $\theta(t)$—we compute reaction forces and moment—$R_1(t)$, $R_2(t)$, and $M(t)$—in the time domain. The displacements, as well as the reaction forces and the moment are, in turn, transformed from the time domain into the frequency domain via FFT. We then compute the impedance functions for the frequency spectrum included in the loading time history. Here, in order to capture the impedance matrix for a broad range of frequencies, we use a Ricker pulse, shown in Fig. 3, to modulate the rigid vertical, horizontal, and rotational movements—i.e., $\Delta_1(t)$, $\Delta_2(t)$, and $\theta(t)$.

(a)   (b)
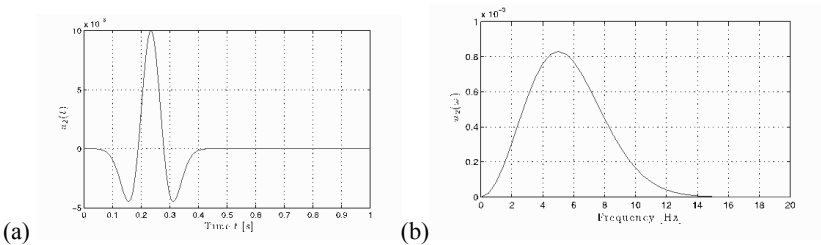
Figure 3. (a) The Ricker pulse loading time-history of the known displacement $u_2(t)$ with (b) the frequency spectrum of the central frequency f_r = 5 Hz.

To compute the time-domain wave responses, we consider a solid medium in a plane-strain setting as shown in Fig. 4. The regular domain (Ω_0) is truncated by

PMLs. In the regular domain Ω_c, the wave motion is governed by the elastic wave equation. In the PML domain ($\Omega \setminus \Omega_c$), a hybrid-PML formulation—developed by Kuckoban and Kallivokas (2011)—gives rise to the attenuation of the amplitudes of both propagating and evanescent waves. The two wave equations are coupled with the proper interface conditions—*viz.*, displacement and traction continuity conditions.
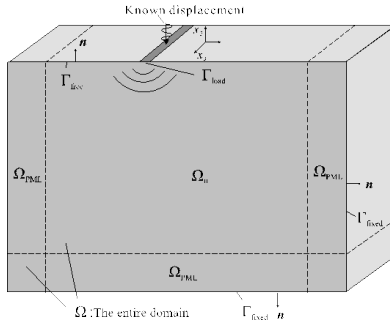
Figure 4. The two-dimensional wave model for a semi-infinite soil medium.

The wave response solutions are obtained via the standard finite element method, which is implemented on a high-performance computing platform. The computed wave responses and impedance functions are rigorously verified by comparing them with reference solutions (Luco and Westmann, 1972; Rajapakse and Shah, 1988). Further formulation details and the verification studies are omitted here for brevity.

NUMERICAL EXPERIMENTS

In this section, we discuss the variation of the impedance functions of rigid massless foundations for three different configurations (surface, embedded shape, and rectangular void), shown in Fig. 5, with respect to variations in the soil profile. For brevity, we consider only horizontally layered or linearly-depth-dependent soils here.
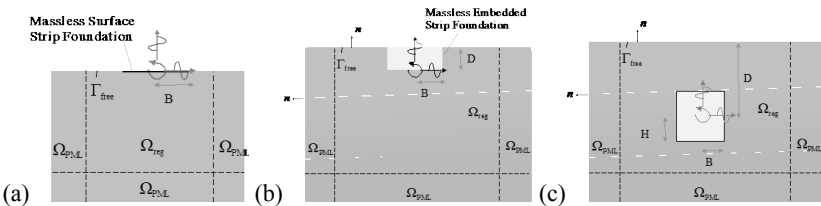
Figure 5. Rigid foundations of different types in a 2D setting: (a) strip surface foundation, (b) embedded foundation, and (c) rectangular void (tunnel).

Surface Foundation Resting on A Two-Layered Soil Medium: First, we

consider two-layered soil media. The Young's moduli of the top layer and the underlying half-space are respectively denoted by E_1 and E_2. We consider the stiffness ratio E_1/E_2 of 1.5, 2, and 3. The Poisson's ratio and the mass density are uniform for the entire soil domain. The interface between the two layers is located at $x_2 = -3.75B$, where B denotes the half-width of the surface strip foundation. We use a Ricker pulse with non-dimensional central frequency $a_0 = \omega_{cr}B/V_s = 0.48$. V_s denotes the shear wave velocity of the top soil layer.
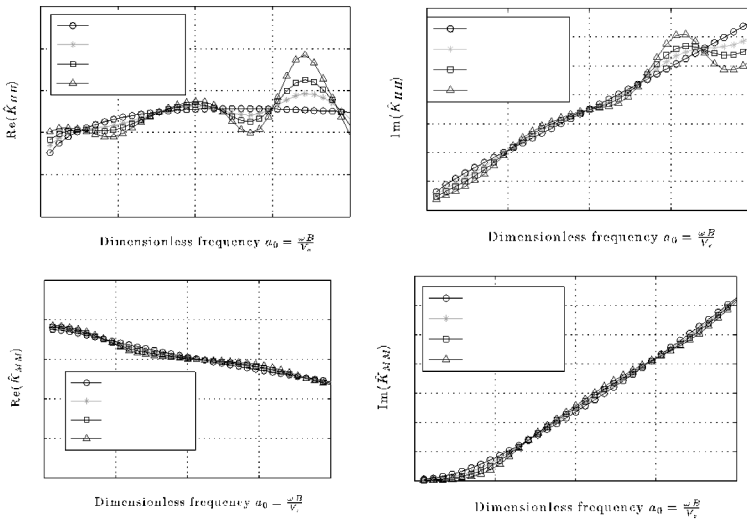
Figure 6. The impedance functions (K_{HH} and K_{MM}) of the strip surface foundation on a two-layered soil medium.

As seen from Fig. 6, the impedance functions become more oscillatory as the stiffness contrast becomes more pronounced. The waviness occurs because of wave reflections between the ground surface and the layers' interface. That is, for some frequencies, constructive interference occurs, which amplifies the reaction response at the foundation boundary (i.e., the real part of K increases at the given frequency), whereas for some other frequencies the interference is destructive (Dobry et al., 1976; Jeong et al., 2010)

Rectangular Void (Tunnel) Foundation Embedded in Soil with a Linearly Depth-Dependent Stiffness: A soil profile with a linearly-depth-dependent Young's modulus is considered as shown in Fig. 7, where E_i denotes the Young's modulus at $x_{2(i)}$, which is the depth of the i-th point of inflection of the gradient of the linearly varying modulus. In this numerical experiment, we used the stiffness ratio ($E_2 = E_3)/E_1$ of 0.5, 2, 3, and 4. The parameter $x_{2(1)}$ is 0 (the top ground surface); $x_{2(2)}$ is $-1.78D$; and $x_{2(3)}$ is $-2.67D$. The Poisson's ratio and the mass density are uniform for

the entire soil domain.

We consider a rectangular void with a half-width of B and a half height of H (cf. Fig. 5), where $B=H$ is considered. The depth from the ground surface to the centroid of the void is $D = 3 \times B$. The location of the rotational centroid is $(x_{1c}, x_{2c}) = (0, -D)$. The Ricker pulse loadings with the non-dimensional central frequency $a_0 = (\omega_{cr} \times B)/V_s = 0.9$ are used. Fig. 8 indicates that both the real and the imaginary parts of K_V, K_{HH}, K_{MM}, and K_{HM} for this rectangular void foundation are highly sensitive to the gradient of the soil's Young's modulus.
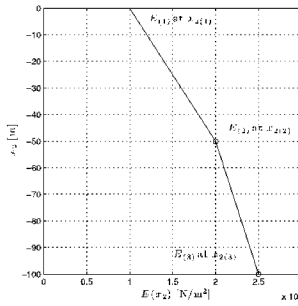
Figure 7. Linearly-depth-dependent Young's modulus $E(x_2)$.
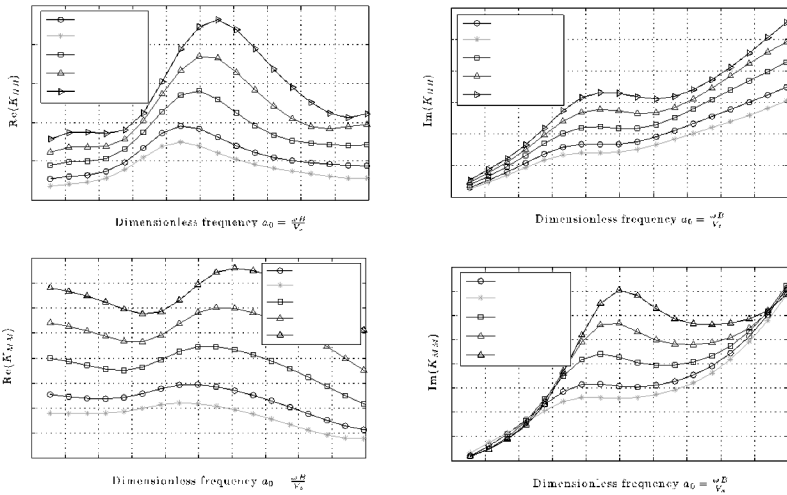
Figure 8. The impedance functions (K_{HH} and K_{MM}) of the rectangular void foundation within soil of linearly-varying soil property.

5. CONCLUSIONS

We presented a numerical approach to computing impedance functions for rigid foundations with arbitrary shapes that are resting on (or embedded in) heterogeneous soil media. PMLs are used for truncating the semi-infinite domain. The results indicate that the impedance functions are quite sensitive to the foundation type and geometry, as well as soil heterogeneity. This result suggests that impedance functions should be systematically investigated for various soil profiles and foundation types so that they can be scaled and utilized for the seismic design of structures for which SSI is important.

In future work, we plan to extend our method to the computation of impedance functions for flexible foundations, wherein an impedance function matrix that corresponds to all of the degrees of freedom on the flexible foundation boundary is computed. Alternatively, a reduced set of vibrational modes of the flexible foundation boundary can be selected and impedance functions can be computed for each of those modes.

6. REFERENCES

Bielak, J. and Christiano, P. (1985). "On the effective seismic input for non-linear soil-structure-interaction systems." *Earthquake Engineering & Structural Dynamics*, 12, 107–119.

Dobry, R., Oweis, I., and Urzua, A. (1976). "Simplified procedures for estimating the fundamental period of a soil profile." *Computer Methods in Applied Mechanics & Engineering*, 192, 1337–1375.

Gazetas, G. (1984) "Seismic response of end-bearing single piles." *Soil Dynamics and Earthquake Engineering*, 3, 82–93.

Jeong, C., Kallivokas, L. F., Huh, C., and Lake, L.W. (2010). "Optimization of sources for focusing wave energy in targeted formations." *Journal of Geophysics & Engineering*, 7, 242–256.

Kucukcoban, S. and Kallivokas, L. F. (2011). "Mixed perfectly-matched-layers for direct transient analysis in 2D elastic heterogeneous media." *Computer Methods in Applied Mechanics and Engineering*, 200, 57–76.

Luco, J. E. and Westmann, R. A. (1972). "Dynamic response of a rigid footing bonded to an elastic half space." *Journal of Applied Mechanics,* 39, 527–534**.**

Mylonakis, G., Nikolaou, A., and Gazetas, G. (1997). "Soil-pile-bridge seismic interaction: kinematic and inertial effects. Part I: soft soil." *Earthquake Engineering & Structural Dynamics*, 26, 337–359.

Rajapakse, R. K. N. D. and Shah, A. H. (1988) "Impedance of embedded rigid strip foundations." *Earthquake Engineering & Structural Dynamics,* 16,255–273.

Safak, E. (2006). "Time-domain representation of frequency-dependent foundation impedance functions." *Soil Dynamics & Earthquake Engineering*, 26, 65–70.

Wolf, J. P. (1985). "*Dynamic soil-structure interaction*." Prentice Hall.

Wolf, J. P. (1989). "Soil-structure-interaction analysis in time domain." *Nuclear Engineering & Design*, 111, 381–393.

Efficient Analysis and Optimization of Reconstruction Plans for Damaged Transportation Networks Following Disasters

Omar El-Anwar[1], Jin Ye[2], and Wallied Orabi[3]

[1]Department of Construction Management, University of Washington, Seattle, WA 98195; PH (206) 543-4736; FAX (206) 685-1976; email: elanwar@uw.edu
[2]Department of Industrial and Systems Engineering, University of Washington, Seattle, WA 98195; email: jinye@uw.edu
[3]OHL School of Construction, Florida International University, Miami, FL 33171; PH (305) 348-2730; FAX (305) 348-6255; email: worabi@fiu.edu

ABSTRACT

Planning for post-disaster reconstruction of damaged transportation networks is a complex and computationally expensive effort. Decision makers need to prioritize the use of limited resources to maximize societal benefit through accelerated traffic services restoration, which involves evaluating numerous recovery scenarios. This paper presents a new model that enables efficient analysis and optimization of the post-disaster reconstruction of transportation networks that is capable of identifying optimal reconstruction plans using reasonable computational overhead. The model employs a number of search space reduction techniques using mixed linear-integer programming and goal programming inspired approaches in order to explore the large space of possible alternative solutions. Each of these solutions is comprised of a set of interdependent variables, including (1) contractors selection; and (2) the start dates of reconstruction projects and their constituent work packages. An application example is used to demonstrate the new model's superior computational performance compared to recent research developments.

INTRODUCTION

Planning the restoration of transportation services following disasters is a complex and challenging task to planners and decision makers mainly due to the difficulty of searching for and implementing an optimal reconstruction plan that satisfies conflicting planning objectives (Orabi et al. 2009). These plans should prioritize the reconstruction efforts and assign them to qualified contractors in such a way that simultaneously minimizes the disruption in transportation services and reconstruction costs. Several studies focused on restoration of damaged transportation networks following natural disasters from different perspectives. First, some studies focused on optimizing the reconstruction efforts of damaged transportation networks in order to maximize network performance, minimize reconstruction costs, or minimize reconstruction duration (Chen and Tzeng 1999; Orabi et al. 2009). Second, other studies investigated assessing and measuring the performance of transportation network following disasters (Chang and Nojima 2001; Chen and Eguchi 2003; Nojima and Sugito 2000). Third, analyzing policies and strategic planning for mitigating the impacts of disasters were the focus of other research studies (Chang 2003; Gunes and Kovel 2000; Opricovic and Tzeng 2002).

Despite the contributions of these original research studies, no reported research was able to optimize the post-disaster reconstruction problem of transportation networks at reasonable and practical computational effort. Therefore, there is a need for a new model that is capable of searching for and identifying optimal reconstruction plan(s) at reasonable computational cost.

PROBLEM STATEMENT

The objective of this paper is to present an efficient analysis and optimization model capable of assisting decision makers in (1) prioritizing reconstruction projects and identifying their optimal start dates; and (2) determining the optimal contractor assignments. Previous studies reported the computational complexity of this planning problem, which is mainly attributed to the computationally expensive transportation network analysis combined with a huge search space (Orabi 2010). For instance, for a post-disaster reconstruction program that involves seven reconstruction projects and three competing contractors, there are 5,040 (i.e. 7!) possible prioritization scenarios that need to be analyzed. Furthermore, for each prioritization scenario, there are 2,187 (i.e. 3^7) possible contractor assignments. Accordingly, decision makers need to identify the optimal solution out of over 11 million ($5,040 \times 2,187$) alternatives for this reconstruction plan. The huge computational cost of this problem is even aggravated due to the need to perform a transportation network analysis for each of these alternatives in order to evaluate its traffic performance. Solving a problem of this size can take up to few weeks of continuous running time, and this computational cost increases exponentially with the size of the transportation network analyzed (Bell and Iida 1997). In this paper, a problem of the same size (seven projects and three contractors) is utilized throughout the paper to demonstrate the solution presented to overcome the huge computational effort.

To overcome the aforementioned computational challenges, this paper presents a new model that incorporates four innovative decomposition and search space reduction techniques to efficiently optimize the reconstruction planning problem. The application of these techniques requires inverting the prioritization of reconstruction projects to use their proposed finish dates instead of the traditional way of using start dates. For example, in this paper, (5,7,3,4,1,2,6) denotes a prioritization scenario where project #5 would be the first project to finish, followed by project #7 and so on until project #6 finishes last. The start dates of these projects can then be derived backwards according to the selected contractors and their resources availability and productivity rates. The following sections of the paper describe four decomposition and search space reduction techniques that comprise the new model; namely (i) traffic decomposition; (ii) branch elimination; (iii) scenario elimination; and (iv) constrained search.

I. TRAFFIC DECOMPOSITION

This innovative strategy eliminates the need to run transportation network analysis for each of the 11 million possible planning alternatives. It should be noted that the proposed model utilizes deterministic user equilibrium (DUE) to determine the traffic flow on each network link. DUE provides an acceptable estimation of traffic flow at a reasonable computational cost compared to stochastic user

equilibrium (Bell and Iida 1997). Each possible reconstruction plan (comprising of a prioritization scenario and an associated contractor assignment) has a unique impact on the traffic restoration rate. However, leveraging the inverted approach, all these plans can be decomposed into a considerable smaller set of incremental changes in traffic performance. Each incremental change occurs when one of the reconstruction projects is completed and is open for the public, while the magnitude of the change itself is a function in (1) the completed project and (2) the state of network at the time of its completion (i.e. which other reconstruction projects have already been completed). In the case of seven projects, there are seven incremental changes to the transportation performance at the completion of each project. At each incremental change, there is a set of possible values, which can be computed as $^{N}C_{n}$, where N is the total number of projects (e.g. seven projects) and n is the total number of completed projects. Figure 1 shows the comparison between enumerating and analyzing each of the 11 million plans using DUE versus decomposing the plans into their constituent incremental changes. Using traffic decomposition, there will only be 127 incremental changes (which are equal to $2^{7}-1$), where each incremental change can be evaluated using DUE. These 127 incremental changes are then stored in a database in order to evaluate the traffic performance of each specific contractor assignment under each prioritization scenario without the need to run additional transportation network analysis. Compared to the initial need to run DUE 11 million times, this is a tremendous saving of 99.999% in computational effort.
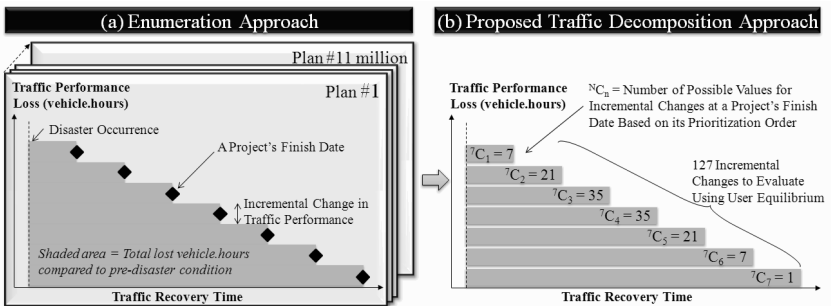
Figure 1. Computational Savings Using Traffic Decomposition.

II. BRANCH ELIMINATION

As shown in Figure 2, this problem's search space can be visualized as branches of projects to complete, where the full length of each branch represents the order of projects finish dates in a unique prioritization scenario. For seven projects, there are seven major branches, each starting with completing one of the seven projects. If any of the branches could be eliminated, it automatically cuts one seventh (i.e. 1/N) of the search space (about 1.6 million solutions).
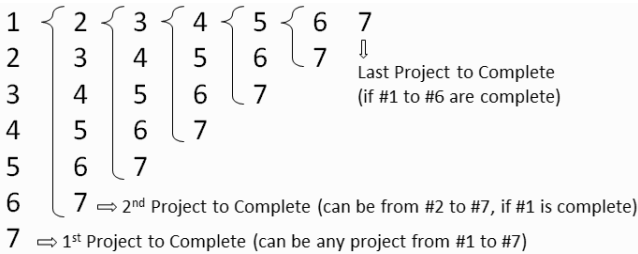
Figure 2. Possible Prioritization Scenarios Forming Branches of Solutions.

The branch elimination strategy consists of two steps: (1) creating a baseline solution with a promising performance using heuristics and (2) eliminating branches of solutions that are guaranteed to generate inferior performance compared to the baseline solution both in terms of construction costs and traffic restoration rate. The details of the heuristics involved in creating the baseline solution are beyond the scope of this paper; but in essence, they utilize series of mixed linear integer programming (MLIP) runs to generate a solution that requires (1) the least possible construction costs and (2) an enhanced traffic restoration rate (although not optimal).

The second step compares the traffic performance of the baseline solution to the potential performance of each branch of solutions. Since the baseline solution already guarantees minimum construction costs, then the comparison is based on traffic performance only. To this end, the model first computes the traffic performance loss in vehicle.hours (veh.hrs) for the baseline solution ($TL^{baseline}$, which is equal to the shaded area in Figure 3a). The model then computes a target completion date (TCT) that the first project must be completed by in any proposed prioritization scenario; otherwise, that prioritization scenario will result in higher traffic losses than $TL^{baseline}$. Accordingly, a branch of solutions that starts with a project whose duration is greater than TCT will definitely result in inferior solutions compared to the baseline solution. As such, this whole branch of solutions can be eliminated from the solutions space. As previously mentioned, in the case of seven projects, eliminating one branch of solutions cuts about 1.6 million solutions from the solution space. Furthermore, similar investigations can be applied on sub-branches. For example, for a branch whose first project duration is less than TCT, another target completion date can computed for its second project, and so on.
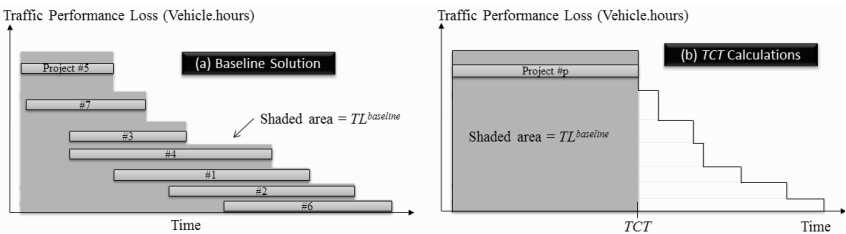
Figure 3. Establishing a Target Completion Time (TCT)

III. SCENARIO ELIMINATION

While the previous eliminations strategy focused on eliminating branches or sub-branches of solutions, this strategy focuses on eliminating specific prioritization scenarios. It compares the potential traffic performance of each scenario to that of baseline solution ($TL^{baseline}$). This strategy first computes a minimum bound to the traffic losses (TL^{min_bound}) a prioritization scenario can result in. This is computed by identifying the contractor assignments that would minimize each project's duration while assuming that the contractors have unlimited resources (i.e. no projects will be delayed because of other projects competing on the same pool of resources). Secondly, the model computes the traffic losses under these assumptions while complying with the projects' completion order dictated by the considered prioritization scenario. Hence, the computed traffic losses represent the lower bound (TL^{min_bound}). Since the baseline solution is a feasible solution that guarantees the least construction costs, then if TL^{min_bound} is larger than or equal to $TL^{baseline}$, then all solution under that prioritization scenario are inferior to the baseline solution and that scenario can be safely eliminated. Eliminating each prioritization scenario for a case of three contractors competing on seven projects will cut the solution space by 2,187 solutions.

IV. CONSTRAINED SEARCH

The earlier mentioned inverted approach of scheduling projects using their finish dates (rather than start dates) enables the use of MLIP to identify the optimal contractor assignments for a given prioritization scenario within reasonable computational time. However, there is still the challenge of identifying the optimal prioritization scenario to start with. Hence, after reducing the search space using the branch and scenario elimination strategies, a constrained search strategy is developed to identify the global optimal reconstruction plans using a goal programming inspired approach.

The overall objective of the proposed optimization model is to identify the optimal tradeoffs between minimizing traffic losses (*TL*) and total construction costs (*TCC*). This strategy applies varying steps to identify different global optimal solutions, as shown in Figure 4.
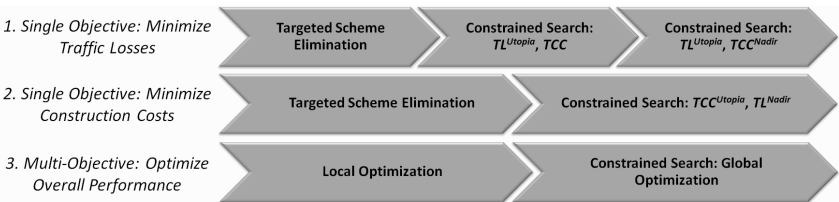
Figure 4. Constrained Search Strategies

1. Minimize Traffic Losses

The model follows three steps to identify the optimal reconstruction plan that achieves the least traffic losses (TL^{Utopia}) and its associated construction costs (TCC^{Nadir}). The model first applies the scenario elimination strategy with the target of eliminating any prioritization scenario that would not outperform a baseline solution with improved traffic performance. These eliminations focus only on traffic performance with no regard to construction cost since the objective is to minimize traffic losses only.

Second, the model utilizes MLIP to identify a feasible solution that can achieve the least traffic losses (TL^{Utopia}). To this end, the model runs MLIP with each remaining prioritization scenario in the solution space with the objective of identifying one plan under each scenario that would minimize TL. However, the number of scenarios to investigate can be in the order of hundreds or even thousands. To reduce the computational time of this step, the model applies a constrained search strategy for each MLIP run. This strategy keeps track of the least identified TL among all investigated scenarios and adds a constraint that each scenario under investigation should not result in a solution that can exceed that TL. This added constraint deems any inferior solution from the traffic losses standpoint infeasible. Accordingly, the model does not consume computational time trying to optimize a plan that will eventually result in an inferior solution, but rather eliminates the whole plan from the solution space. This strategy has proven to save significant running time. This step results in generating a plan that can achieve TL^{Utopia}, but with associated construction costs that can still be improved (i.e. TCC is not guaranteed to have reached TCC^{Nadir}).

The third step identifies the global optimal plan that achieves both TL^{Utopia} and TCC^{Nadir}. To this end, the model uses MLIP to optimize the contractor assignments for the prioritization scenario that achieves TL^{Utopia} with the objective of minimizing TCC. Since there are can be thousands of possible assignments within that scenario, the model applies the constrained search strategy but this time with two added constraints: (1) the generated plan's traffic losses should not exceed TL^{Utopia}; and (2) the plan's TCC should not exceed the least identified costs for solutions in that prioritization scenario (where the model keeps track of the least identified TCC within the investigated scenario). As such, the model eventually identifies the plan that achieves both TL^{Utopia} and TCC^{Nadir}.

2. Minimize Construction Costs

As shown in Figure 4, the model follows two steps to identify the optimal reconstruction plan that achieves the least construction costs (TCC^{Utopia}) and its associated traffic losses (TL^{Nadir}). Similar to the previous section, the model first starts by reducing the search space using a targeted scenario eliminations strategy. In the second step, the model identifies the global optimal plan that achieves both TCC^{Utopia} and TL^{Nadir} by applying MLIP to each of the remaining prioritization scenarios. Since any scenario can minimize costs by awarding the lowest bidders, MLIP is used with the objective of minimizing TL (and not TCC), while constrained search is used to ensure achieving TCC^{Utopia} and TL^{Nadir} with the least computational

resources. In this case, the model keeps track of the least *TL* while constraining the generated plans to not exceed the least identified *TL* as well as TCC^{Utopia} (which is already a known value). Hence, the model identifies the second global optimal plan on the other end of the Pareto front, which achieves TCC^{Utopia} and TL^{Nadir}.

3. Optimize Overall Performance

In order to generate any other global optimal plan, the optimization objective should include a weighted combination of *TL* and *TCC*. In such a multi-objective setting, the overall performance of any generated plan can be computing using Equation (1).

$$OP = w_{TL} \times \left(\frac{TL^{Nadir} - TL}{TL^{Nadir} - TL^{Utopia}} \right) + w_{TCC} \times \left(\frac{TCC^{Nadir} - TCC}{TCC^{Nadir} - TCC^{Utopia}} \right) \qquad (1)$$

Where, *OP* is the overall performance of a reconstruction plan; w_{TL} and w_{TCC} are the relative weights of each of minimizing traffic losses and reconstruction costs, respectively; *TL* is the total traffic losses for the candidate reconstruction plan measured in added veh.hrs; TL^{Nadir} and TL^{Utopia} are computed in the previous steps and represent the highest and least traffic losses among Pareto optimal solutions, respectively; *TCC* represents the total construction costs for the candidate reconstruction plan; and TCC^{Nadir} and TCC^{Utopia} are also already computed and represent the highest and least total reconstruction costs among Pareto optimal solutions, respectively. *OP* can range between 0 (lowest performance) and 1 (highest performance).

For any unique combination of objectives relative weights (w_{TL} and w_{TCC}), the model employs two steps to generate the corresponding global optimal plan. The model first focuses on the prioritization scenario that generated the initial baseline solution and utilizes MLIP to generate a local optimal solution within that scenario with the objective of maximizing *OP*. In the second step, the model utilizes the local optimal solution as a starting point to generate a global optimal solution that corresponds to the unique combination of relative weights. This is achieved by applying MLIP to the remaining prioritization scenarios in the solution space in order to maximize *OP*. Constrained search is employed by keeping track of the highest *OP* and requiring any generated solution to exceed that performance. Accordingly, the model eventually identifies the required global optimal plan. These two steps can be repeated for any required combination of objectives relative weights. These combinations can be pre-defined by decision makers or selected during the optimization process based on the results of previous runs.

APPLICATION EXAMPLE

In order to demonstrate the efficiency of the newly developed model, an application example from a previous research study (Orabi et al. 2009) is analyzed to show the huge savings in the computational cost as a result of the aforementioned decomposition and space reduction techniques. This example involves searching for optimal reconstruction plan(s) for seven transportation projects and three competing contractors. Therefore, the search space for this problem includes over 11 million

solutions. In previous research, the optimal/near-optimal solutions were generated for this problem using a multi-objective genetic algorithm model after several days of continuous running time on a dual core personal computer (Orabi et al. 2009).

The proposed model was used to optimize a problem of the same size and scale. First, traffic decomposition was used to decompose the traffic analysis problem into 127 distinct cases of incremental traffic improvements, where each case was evaluated using DUE. This data was then stored in the traffic analysis database for retrieval during the optimization process. Second, the model applied the branch elimination strategy, which could eliminate two branches of solutions out of seven existing branches. This strategy eliminated all prioritization scenarios that require finishing project #2 or #6 first, cutting the search space by about 3.15 million solutions. Third, the scenario elimination strategy was employed to the remaining solutions. This strategy further eliminated 1,080 inferior prioritization scenarios, cutting the search space by about 2.36 million solutions.

The combined effect of these elimination strategies resulted in a reduced search space of about 5.5 million possible solutions (50% reduction compared to the original 11 million solutions). Constrained search was then applied to this reduced search space. As such, the model identified (TL^{Utopia}, TCC^{Nadir}) at (736,139,824 veh.hrs, $19,818,894) and ($TL^{Nadir}$, TCC^{Utopia}) at (902,609,624 veh.hrs, $19,142,253). Furthermore, the weight combination (w_{TL}, w_{TCC}) of (0.5, 0.5) was applied and resulted in an optimal solution with (TL, TCC) at (740,645,773 veh.hrs, $19,381,380). The total running time for this model to optimize the aforementioned problem was less than 11 hours on a consumer-grade computer with 4.5 hours of average running time needed to generate any additional optimal solution at a different relative weight combination (compared to the original 4-5 days of running time for the multi-objective genetic algorithm model developed by Orabi et al. 2009).

CONCLUSIONS AND FUTURE WORK

This paper presented the development of new optimization model to address the highly complex problem of optimizing post-disaster transportation reconstruction plans. The proposed model aims at identifying the optimal start dates and contractor assignments for the reconstruction projects that would accelerate traffic restoration while minimizing total construction costs. This model employs four traffic decomposition and solution space reduction techniques which significantly reduce the computational requirements for this optimization problem while maintaining the quality of generated solutions. The first strategy applies an innovative traffic decomposition technique that eliminates almost all of the traffic analysis effort. The second strategy eliminated complete branches and sub-branches of solutions using conservative traffic analysis and computations. The third strategy identifies and eliminates inferior prioritization scenarios by computing their lower bound traffic losses. While, the fourth strategy adopts a constrained search approach to significantly reduce the computational time needed to generate each global optimal solution. An application example is presented to demonstrate the superior computational performance of the new model. It should, however, be noted that this model utilizes mixed linear integer programming which reduces the modeling flexibility of the optimization problem (compared to genetic algorithms) because of its strict linear formulation requirements. While, this limitation did not impact

optimizing the problem in hand, it should be taken into consideration when applying these strategies to other problems.

ACKNOWLEDGEMENTS

REFERENCES

Bell, M. G. H., and Iida, Y. (1997. Transportation network analysis, Wiley, Chichester, U.K.

Chang, S. E. (2003). "Evaluating disaster mitigations: Methodology for urban infrastructure systems." *Natural Hazards Review*, 4(4), 186–196.

Chang, S. E., and Nojima, N. (2001). "Measuring post-disaster transportation system performance: The 1995 Kobe earthquake in comparative perspective." *Transportation Research, Part A: Policy Practices*, 35, 475–494.

Chen, Y., and Eguchi, R. T. (2003). "Post-earthquake road unblocked reliability estimation based on an analysis of randomicity of traffic demands and road capacities." *Proceedings of the 6th U.S. Conference and Workshop on Lifeline Earthquake Engineering*, Long Beach, CA, ASCE, Reston, VA, 93–102.

Chen, Y.-W., and Tzeng, G.-H. (1999). "A fuzzy multi-objective model for reconstructing the post-quake road-network by genetic algorithm." *International Journal of Fuzzy Systems*, 1(2), 85–95.

Gunes, A. E., and Kovel, J. P. (2000). "Using GIS in emergency management operations." *Journal of Urban Planning and Development*, 126(3), 136–149.

Nojima, N., and Sugito, M. (2000). "Simulation and evaluation of postearthquake functional performance of transportation network." *Proceedings of the 12th World Conference on Earthquake Engineering*, NZ Society of Earthquake Engineering, Auckland, New Zealand.

Opricovic, S., and Tzeng, G.-H. (2002). "Multicriteria planning of postearthquake sustainable reconstruction." *Computer Aided Civil Infrastructure Engineering*, 17(3), 211–220.

Orabi, W. (2010). "Optimizing highway reconstruction and rehabilitation projects." Doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL.

Orabi, W., El-Rayes, K., Senouci, A. B., and Al-Derham, H. (2009). "Optimizing postdisaster reconstruction planning for damaged transportation networks." *Journal of Construction Engineering and Management*, 135(10), 1039–1048.

Comparing image features and machine learning algorithms for real-time parking space classification

M. Tschentscher[1], M. Neuhausen[1], C. Koch[2], M. König[2], J. Salmen[1], M. Schlipsing[1]

[1]Institute for Neural Computing, Ruhr-Universität Bochum,
Universitätsstraße 150, 44780 Bochum, Germany;
PH (+49) 234-32 27 977; FAX (+49) 234-32 14 210;
email: marc-philipp.tschentscher@ini.rub.de
[2]Chair of Computing in Engineering, Ruhr-Universität Bochum,
Universitätsstraße 150, 44780 Bochum, Germany;
PH (+49) 234-32 26 174; FAX (+49) 234-32 14 292;
email: koch@inf.bi.rub.de

ABSTRACT

Finding a vacant parking lot in urban areas is mostly time-consuming and not satisfying for potential visitors or customers. Efficient car-park routing systems could support drivers to find an unoccupied parking lot. Current systems detecting vacant parking lots are either very expensive due to the hardware requirement or do not provide a detailed occupancy map. In this paper, we propose a video-based system for low-cost parking space classification. A wide-angle lens camera is used in combination with a desktop computer. We evaluate image features and machine learning algorithms to determine the occupancy of parking lots. Each combination of feature set and classifier was trained and tested on our dataset containing approximately 10,000 samples. We assessed the performance of all combinations of feature extraction and classification methods. Our final system, incorporating temporal filtering, reached an accuracy of 99.8 %.

INTRODUCTION

In urban areas, finding a vacant parking lot in parking garages or parking lots is time-consuming and a tedious task for drivers. A system to detect available parking spaces to route drivers efficiently to proper lots is desirable. Some systems have reached the market or are under research promising to support the driver by locating a vacant parking lot. The oncoming section gives a detailed overview. A video-based system offers a proper alternative to deal with the classification problem. It is possible to combine low-cost hardware requirements with providing detailed occupancy maps for parking areas, which most of the current systems do not provide. As we show in this paper, several image processing and machine learning algorithms which can be employed to classify parking lots already exist. By using video-based systems several challenges occur especially on outdoor car-parks. Different weather and lighting conditions or objects occluding parking lots might influence the accuracy for the given task. In the implementation and the results section our experiments and the best

performing feature combination and machine learning algorithm are shown. Finally, a conclusion about the developed system is given.

RELATED WORK

Sensor-based methods. A common type of pavement embedded sensors are inductive loop detectors (ILDs), which are wire loops either installed at the car park entrance to count entering and leaving vehicles or at each parking space leading to expensive and disruptive maintenance work (Ristola 1992). Another type of embedded systems uses magnetic field sensors, that measures changes in the magnetic flux to detect parking vehicles (Wolff et al. 2006). These kind of sensors need to be employed at each parking lot which can be very costly as each sensing unit is usually attached with a processing unit and a transceiver (Bong et al. 2008). Overhead occupancy detection methods are either based on light, sound or radar sensor systems. The drawback of infrared sensors is their sensitivity towards environmental conditions such as heavy rain, dense fog and blowing snow (Idris et al. 2009). Sound based sensors are insensitive to humidity, but large temperature changes and extreme air turbulence negatively affect their performance. Radar sensors perform well in inclement weather conditions, but sometimes need to be equipped with auxiliary sensors to detect stopped vehicles (Ichihashi et al. 2010). In general, overhead technologies are difficult to install in large outdoor car parks, which limits their applicability in such environments.

Video-based methods. Video-based systems have gathered great attention in recent years (Huang and Wang 2010), since they have the potential to provide a cost effective solution as they allow wide area detection and regular maintenance is possible without disturbing the traffic flow. Moreover, they can use existing visual surveillance infrastructure such as security cameras to capture images and videos (Nallamuthu and Lokala 2008).

True (2007) has combined car feature point detection and color histogram classification using the k-nearest neighbor algorithm and support vector machines to detect vacant parking spaces. The limitations of this work are the relatively low detection accuracy (94%). Ichihashi et al. (2012) have improved the detection performance using fuzzy c-means (FCM) clustering and hyperparameter tuning by particle swarm optimization (PSO). Their system has reached a detection rate of 99.6% for outdoor environments. However, they have not reported on the method's performance in terms of real-time applicability. Tsai et al. (2007) have presented a general approach to the detection of vehicles on highways, road intersections, and parking lots under different weather conditions and vehicle orientations. In this approach, a cascaded multichannel classifier based on corner features, edge features and wavelet coefficients was trained to verify the vehicle detection. However, the system is solely based on static images and has reached an average accuracy of merely 94.9%.

Other methods take advantage of the homogeneous appearance of vacant parking spaces. For example, Yamada and Mizuno (2001) have proposed a homogeneity measure by calculating the area of fragmental segments. Although their system has reached a detection rate of 98.7% with real-time performance, shadows

and occlusions caused by adjacent cars are ignored. To reduce perspective effects, Lopez-Sastre et al. (2007) have suggested applying a Garbor filter bank on rectified images to derive the homogeneity feature for vacant lot detection. Their method has reached an overall classification rate of 99.7%, but it might fail in cases of strong shadows and over-exposure effects which violate homogeneity assumption. Wu et al. (2007) have presented a multi-class SVM to classify the state of three neighboring spaces as a unit. The reported performance is an error rate of 2.6% with real-time capability.

Huang and Wang (2010) have presented a Bayesian detection framework that takes into account both a ground plane model and a vehicle model to represent environmental changes. Although high accuracy of 99.4% has been reported, their system does not reach real-time performance. Recently, Kabak and Turgut (2010) and Seo and Urmson (2009) have presented methods to detect vacant parking spaces using aerial images. However, real-time applicability has not been in the focus due to the latency in image acquisition. A different approach is to use two cameras. For example, Jung et al. (2005) presented a method based on pixel structure classification and feature-based stereo matching to extract 3D information in real time. However, this approach requires sufficient overlapping of the two camera views, which prevents it from being practical on large car parks.

METHODOLOGY

In this section the methodology of the classification system as proposed in this paper is described. Figure 1 shows an overview of the involved modules.
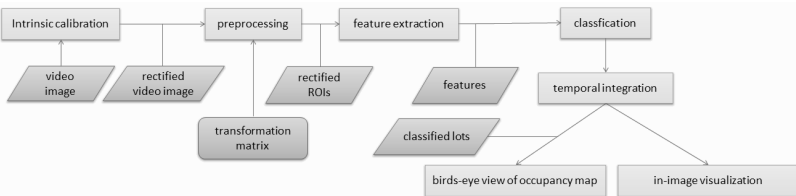
Figure 1. Overview about the different modules used for classification.

Intrinsic Calibration As a first step, one has to develop an intrinsic calibration for the camera. Due to the wide-angle lens of the camera we used in our experiments, the images are radially distorted. To get undistorted images, we used the *radial distortion model* (Weng et al. 1992).

Preprocessing In order to get exact information of the observed lots one has to provide a transformation from camera-image to the world-frame. Therefore, a direct linear transformation (Abdel-Aziz and Karara 1971) is employed which calculates the required transformation matrix. One has to provide at least four world coordinates and their corresponding points in the image manually.

As a next step, the parking lots to be observed have to be marked in the video image manually. It is sufficient to mark the four edges of each lot. These labeled *regions of interests* (ROIs) are extracted from the image and then rectified to fit a

common size. This is a requirement of the classification. Figure 2(a) shows examples of such rectified ROIs.

Feature Extraction We implemented several image features to test their ability to describe vacant and occupied parking lots. We concentrate on four different features, namely color histograms, gradient histograms, difference-of-Gaussian (doG) histograms and finally Haar features and tested them separately and in combination.
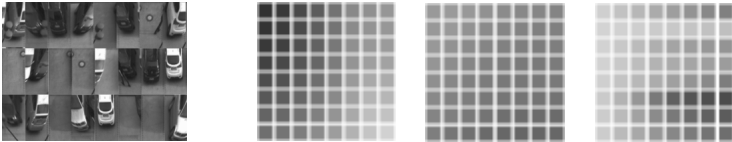
Figure 2(a). Rectified ROIs. (b) Example of RGB, YUV and HSV color spaces.

Color histograms provide detailed information about the distribution of colors in certain images. Therefore we developed an extractor, which is able to create multi-dimensional histograms with different resolutions per channel. We denote (4, 3, 2) as the number of bits for each channel. Figure 2(b) shows an example of the three most popular color spaces (RGB, HSV and YUV). These color spaces are combined with three different edge detection algorithms.

We used a Prewitt operator (Prewitt 1970) that extract edge information in an image and compute a histogram of their distribution (angle histogram). It is possible to define the resolution of these angles by parameter α and defining to use the whole (360) or the semicircle (180) by setting γ to 0 or 1.

The doG method is an alternative filter detecting edges. To extract edge information from a camera image, at first a highly smoothed image (with Gauss-filter) is generated. This image is then subtracted from the original. The histogram represents the distribution of the gray scale values of the resulting image. Figure 3(a) shows thresholded example of doG images.

Haar features (Viola and Jones 2001) are also useful edge-filters which are calculated efficiently with the help of integral images. Haar features can code different orientations of edges. Vertical edges can be coded with features like the first and third whereas the second and fourth detect horizontal edges (Fig. 3(b)).

unoccupied   occupied

Figure 3. (a) DoG images. (b) Example of Haar features

Classification The system has to perform a binary classification (occupied/vacant). Therefor we tested three different learning algorithms such as k-nearest neighbor (k-NN), linear discriminant analysis (LDA) (Hastie and Tibshirani 2001) and support vector machine (SVM) (Vapnik 2008).

In order to further improve the classification results we implemented a temporal integration via *exponential smoothing*. The term $f(t)$ shows the filters equation where α represents the learning rate and Y_t represents the particular class label assigned in time step t for each parking lot. If $\alpha = 1$, no new classification will

affect the older for a specific parking space. In opposite, $\alpha = 0$ means no influence of older results.

$$f(t) = \alpha * Y_{t-1} + (1 - \alpha) * Y_t$$

We build a dataset of approximately 10,000 samples in order to get significant results. To avoid overfitting in the experiments we divided this dataset into two disjoint sets for training and testing.

IMPLEMENTATION AND RESULTS

In this section the implementation of the whole system is explained. Therefore, an overview of the system and the setup is given. The result of each feature combination and machine learning algorithm are shown afterwards.

Hardware and environmental Setup The software runs on a standard desktop PC. In order to cover a possibly large amount of parking spaces, a of-the-shelf wide-angle lens camera is used for recording. It is positioned in the back of the parking lots with a slight top view (see Fig. 4). In our configuration this leads to a monitored area of 15 parking lots per camera. This keeps the costs rather low.

The implemented modules are lined up in a feed-forward pipeline such that the output of a preceding module acts as input of the succeeding. The implemented visualization is shown in Fig. 4.

Figure 4. Camera image and corresponding visualization.

In the following sections, the results of each feature combination are presented. We tested each of the before mentioned classification algorithms with each feature combination. The following tables 1 - 7 show the results (test error) in percent on single images without temporal filtering. The minimal test error is highlighted.

Angle Histogram We denote α as the resolution of the edge direction (1 means every full degree, 36 means every 10th degree is used) and $\gamma = 1$ if using a semicircle as edge direction and $\gamma = 0$ if using the full 360° direction.

We employed k-NN with $k = 5$ and 50 prototypes per class (Tab. 1). The *LDA* was regularized with $\sigma^2 \in [10^{-6}, 1]$ (Tab. 2). The *SVM's parameter* during training are $\gamma \in [0.2, 1]$ and $C \in [10^2, 10^4]$, using a radial basis function as kernel (Tab. 3).

Table 1. Results for angle histogram features in percent for k-NN

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
$\alpha = 1, \gamma = 0$	3.38	3.33	3.65	2.40	2.58	3.53	4.07	3.18	4.10	5.38
$\alpha = 2, \gamma = 0$	3.53	3.77	3.74	**2.38**	2.47	3.36	3.56	3.45	3.92	5.35
$\alpha = 18, \gamma = 0$	3.17	3.59	3.42	3.03	3.15	3.00	4.04	3.86	4.60	5.70
$\alpha = 1, \gamma = 1$	4.27	4.51	3.96	3.00	3.39	3.80	3.95	4.07	6.68	6.39
$\alpha = 2, \gamma = 1$	4.48	4.51	4.99	3.42	3.15	4.07	4.93	4.87	4.16	4.43
$\alpha = 18, \gamma = 1$	4.31	4.90	4.72	4.01	4.78	4.31	5.26	4.48	5.67	6.39

Table 2. Results for angle histogram features in percent for *LDA*

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
$\alpha = 1, \gamma = 0$	7.63	5.20	8.46	6.33	5.02	5.85	8.52	8.26	6.95	11.82
$\alpha = 2, \gamma = 0$	7.22	6.95	9.39	7.16	4.90	6.71	7.87	6.80	6.03	12.15
$\alpha = 18, \gamma = 0$	5.20	3.77	6.50	4.69	4.34	5.44	4.46	4.81	5.14	9.27
$\alpha = 1, \gamma = 1$	6.59	4.93	11.73	6.71	4.48	4.43	7.31	6.65	6.77	10.04
$\alpha = 2, \ \gamma = 1$	5.73	5.32	8.67	6.42	4.31	4.28	6.21	6.21	6.06	9.27
$\alpha = 18, \gamma = 1$	5.17	4.37	5.76	**3.95**	**3.95**	4.81	5.11	4.90	4.34	7.72

Table 3. Results for angle histogram features in percent for *SVM*

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
$\alpha = 1, \gamma = 0$	2.29	2.49	2.73	1.93	**1.75**	2.47	2.72	2.35	2.29	3.50
$\alpha = 2, \gamma = 0$	2.94	2.79	3.18	2.47	1.84	2.14	3.03	3.18	2.88	3.39
$\alpha = 18, \gamma = 0$	3.83	3.03	6.24	3.09	2.94	3.18	3.50	4.48	4.78	5.97
$\alpha = 1, \gamma = 1$	3.24	2.41	2.64	2.14	1.93	2.44	2.97	2.85	3.30	3.03
$\alpha = 2, \ \gamma = 1$	3.00	3.15	2.49	2.20	2.29	2.73	2.85	3.21	3.21	3.53
$\alpha = 18, \gamma = 1$	4.07	2.76	3.89	2.38	2.29	2.55	3.65	3.83	4.43	5.88

DoG Histogram We used four different filter sizes. The classification algorithm is parameterized the same as in angle histograms because feature vectors have similar size. Table 4 – 6 show the results for each used classification algorithm.

Table 4. Results for doG histogram features in percent for *k*-NN

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
3 x 3	3.83	2.76	2.88	3.42	3.03	2.20	3.92	3.09	3.80	5.58
5 x 5	3.39	2.70	2.64	3.47	3.15	2.44	3.50	2.91	3.53	4.96
7 x 7	3.45	2.23	2.23	3.09	**2.14**	3.06	3.86	2.52	3.62	4.31
9 x 9	3.56	2.35	2.70	3.09	**2.14**	2.64	3.68	2.55	3.65	4.51

Table 5. Results for doG histogram features in percent for *LDA*

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
3 x 3	5.20	4.66	4.99	4.13	**3.03**	3.80	5.05	4.46	4.96	8.08
5 x 5	5.38	5.73	4.81	3.83	3.18	4.16	5.20	4.66	4.63	7.34
7 x 7	5.29	5.23	3.98	4.48	4.04	4.34	4.25	4.87	4.63	8.32
9 x 9	4.84	6.27	4.13	4.96	3.83	4.60	5.49	4.01	5.20	9.21

Table 6. Results for doG histogram features in percent for *SVM*

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
3 x 3	3.18	2.55	3.09	3.12	2.47	3.53	3.30	2.97	3.74	4.75
5 x 5	3.06	2.61	1.99	1.96	**1.69**	2.67	3.15	2.49	3.06	3.56
7 x 7	2.82	2.73	2.32	1.96	1.96	3.39	3.33	2.26	3.00	4.10
9 x 9	3.03	2.64	1.72	2.82	2.02	3.24	3.33	2.94	3.00	4.40

Haar Features This section shows the results of Haar features combined with color histograms in Table 7. For parameterizing the SVM, we used different parameters. We set $\gamma \in [0.0003, 0.0015]$ and $C \in [0.01, 1]$.

Table 7. Results for Haar features in percent

	RGB (2,2,2)	RGB (3,3,3)	HSV (4,0,0)	HSV (3,3,0)	HSV (4,4,0)	HSV (3,3,2)	HSV (0,3,3)	YUV (2,2,2)	YUV (3,1,1)	YUV 0,4,4
k-NN	15.74	11.73	12.83	13.15	**10.13**	12.80	14.41	17.02	14.91	18.15
LDA	7.13	6.86	7.01	5.73	**4.25**	5.49	7.81	6.59	6.36	12.00
SVM	5.05	4.69	4.60	4.93	**2.85**	3.74	5.17	4.19	4.40	8.91

It is noteworthy that all of the feature combinations perform best either on HSV(3,3,0) or HSV(4,4,0) color histogram. In preliminary experiments these two color histograms reached the highest accuracy.

Temporal Filtering We employed temporal filtering on whole video sequences in order to improve the results. Varying $\alpha \in [0.7, 0.8]$ and choosing a threshold above 0.8 considering *f(t)*, we achieved classification results of 99.8%.

CONCLUSION

In this paper we developed a promising video-based system for vacant parking space classification using image features and machine learning algorithms. We compared four different image features (color, angle, doG histograms and Haar features) and three different classification algorithms (*k*-NN, LDA, SVM) and showed their performance combining color histograms with one of the others.

The final system relies on color and difference of Gaussian histograms, an SVM classifier and *exponential smoothing* for temporal filtering. This system reached an accuracy of 99.8% and achieved real-time speed.

It is imaginable to use this system in parking garages even though it was only tested on outside parking lots. Because of less space, more cameras are needed to cover a large range of lots. Furthermore the lighting conditions have to be examined and adjusted to get sufficient results. Improvements can be achieved by minimizing the influence of adjacent cars parking left and right from the labeled area. Concerning other car parks or parking garages further experiments should be conducted adapting the system to different lighting conditions.

REFERENCES

Abdel-Aziz, Y. and Karara, H. (1971). Direct linear transformation from comparator coordinates into object-space coordinates in close-range photogrammetry. In *Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry*, pages 1–18. American Society of Photogrammetry.

Bong, D. B. L., Ting, K. C., and Lai, K. C. (2008). Integrated approach in the design of a car park occupancy information system (COINS). *IAENG International Journal of Computer Science*, 35(1):7–14

Hastie, T., Friedman, J., and Tibshirani, R. (2001). *The elements of statistical learning*. Springer Series in Statistics.

Huang, C.-C. and Wang, S.-J. (2010). A hierarchical bayesian generation framework for vacant parking space detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(12):1770–1785.

Ichihashi, H., Katada, T., Fujiyoshi, M., Notsu, A., and Honda, K. (2010). Improvement in the performance of camera based vehicle detector for parking lot. In *Proceedings of the IEEE International Conference on Fuzzy Systems*, pages 1950–1956

Idris, M. Y. I., Leng, Y. Y., Tamil, E. M., Noor, N. M., and Razak, Z. (2009). Car park system: A review of smart parking system and its technology. I*nformation Technology Journal*, 8(2):101–113.

Jung, H. G., Kim, D. S., and Yoon, P. J. und Kim, J. H. (2005). Stereo vision based localization of free parking site. In *Computer Analysis of Images and Patterns*, pages 231–239. Springer

Kabak, M. O. and Turgut, O. (2010). Parking spot detection from aerial images. Final project autumn 2010, machine learning class, Stanford University.

Lopez-Sastre, R. J., Jimenez, P. G., Acevedo, F. J., and Bascon, S. M. (2007). Computer algebra algorithms applied to computer vision in a parking management system. In *Proceedings of the IEEE International Symposium on Industrial Electronics*, pages 1675–1680.

Nallamuthu, A. and Lokala, S. (2008). Vision based parking space classification. Report, Clemson University, Department of ECE.

Prewitt, J.M.S (1970), Object enhancement and extraction, Lipkin, Rosenfeld (Eds.), Picture Processing and Psychopictorics, Academic Press, pages 75–149

Ristola, T. (1992). Parking guidance system in tapiola. In *Proceedings of the IEEE Conference on Road Traffic Monitoring*, page pp. 195.

Seo, Y.-W. and Urmson, C. (2009). Utilizing prior information to enhance selfsupervised aerial image analysis for extracting parking lot structures. In *Proceedingsof the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages339–344.

True, N. (2007). Vacant parking space detection in static images. Report, University of California San Diego.

Tsai, L.-W. and Hsieh, J.-W.and Fan, K.-C. (2007). Vehicle detection using normalized color and edge map. *IEEE Transactions on Image Processing*, 16(3):850–864.

Vapnik, V. N. (2008). *Statistical Learning Theory*. Wiley-Interscience.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 511–518.

Weng, J., Cohen, P., and Herniou, M. (1992). Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on pattern analysis and machine intelligence*, 14(10):965–980.

Wolff, J., Heuer, T., Gao, H., Weinmann, M., Voit, S., and Hartmann, U. (2006). Parkingmonitor system based onmagnetic field sensors. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pages 1275–1279

Wu, Q., Huang, C., Wang, S.-Y., Chiu, W.-C., and Chen, T. (2007). Robust parking space detection considering inter-space correlation. In *Proceedings of the IEEE International Conference on Multimedia and Expo*, pages 659–662

Yamada, K. and Mizuno, M. (2001). A vehicle parking detection method using image segmentation. *Electronics and Communication*, 84(10):25–34.

Computational GIS and Agent-Based Model Development for Routing Optimization to Facilitate Pavement Condition Data Collection

Natalia M. Sanabria[1], Elmira Kalhor[2], Vanessa Valentin[3], Susan M. Bogus[4], Guohui Zhang[5]

[1]MSCE Student and Graduate Research Assistant, Department of Civil Engineering, University of New Mexico, nsanabria@unm.edu
[2]Ph.D Student and Graduate Research Assistant, Department of Civil Engineering, University of New Mexico, ekalhor@unm.edu
[3]Assistant Professor, Department of Civil Engineering, University of New Mexico, vv@unm.edu
[4]Associate Professor, Department of Civil Engineering, University of New Mexico, sbogus@unm.edu
[5]Assistant Professor, Department of Civil Engineering, University of New Mexico, guohui@unm.edu

ABSTRACT

Transportation agencies across the world have been implementing different types of Pavement Management Systems (PMS) to facilitate pavement management and maintenance. Obtaining the necessary information in a manual pavement distress data collection process is costly and time consuming. Research on the data collection of pavement conditions has focused mainly on evaluating the quality of the data collected and there is limited research on ways to improve the efficiency of the data collection process (e.g., selection of optimized routes for collecting data). In this study, Geographical Information Systems (GIS) and Agent Based Modeling (ABM) using swarm intelligence are integrated to optimize decision making processes for optimal routing selection in manual data collection process. The GIS platform provides the road spatial database containing the pavement sections that will be evaluated. Ant Colony Optimization (ACO) was selected as the algorithm to optimize the decision making process of pavement evaluation crews (agents) during the pavement distress data collection process. Both the GIS and ABM models are developed using available data from the manual pavement distress data collection process in summer 2012 for the State of New Mexico. The results of this research will improve the understanding of route selection for pavement data collection.

INTRODUCTION AND BACKGROUND

Transportation agencies across the world have been implementing different types of Pavement Management Systems (PMS) to help with the decision-making processes of designing new roads and providing maintenance to existing ones. There are two types of data collected for a PMS: (1) Inventory data, which describes the physical elements of the road system that do not change markedly over time (e.g., pavement structure, geometrics, costs, and environment) and (2) condition data which describes

the condition of element that can be expected to change over time (e.g., surface distresses, friction, and/or structural capacity) (Bennet 2008; Flintsch et al. 2009). There are two major methods for obtaining the aforementioned data: (1) manual surveys and (2) automated techniques (Bennet 2008; Bogus et al. 2010; Findley et al. 2011; McGhee et al. 2004; Wenjing et al. 2012).

According to Flintsch et al. 2009, *"the collection of network-level pavement condition data, especially pavement distress data, is one of the most costly parts of operating a PMS."* As a result, most Departments of Transportations have opted to outsource the data collection surveys. The increased demand of timely quality data, the reduction in the public sector staff and the availability of more sophisticated equipment (but costly and complex to operate), have been identified as three main factors for this outsourcing tendency (Morian et al. 2002).

Research in this area has focused mainly on evaluating the quality of the data collected, and automated data collection issues such as technology and trade-offs with data collected manually (Bennet 2008; Bennett 2008; Bogus et al. 2010; Flintsch et al. 2009; Morian et al. 2002). There is limited research on ways to improve the efficiency of the data collection process with the available manpower (e.g., selection of optimized routes for collecting data) which could help lower the costs implicated with the both manual and automated surveys.

For a PMS to be useful and due to federal requirements to state agencies in the United States, pavement evaluation data has to be collected periodically (Bennett, C., 2006). Given the high cost of data collection, different means to achieve a cost-efficient process must be created. Optimizing the route planning and selection can result in cost savings such as those obtained in other areas of specialties i.e. parcels deliveries (Pajunas et al. 2007), waste management collection services (Ladanyi 2011; Lim 2010; Sahoo et al. 2005), transportations logistics (Feng et al. 2010; Haughton 2002) and network mobility (Kong et al. 2012; Mosa et al. 2012). Research related to PMS has focused on how to select the appropriate pavement assessment technology to minimize cost (Bennet 2008), but in the area of pavement data collection logistics, research is limited.

METHODOLOGY

GIS and ABM swarm intelligence (SI), were integrated to solve the Pavement Data Collection Routing Optimization (PDCRO) problem. The GIS provided the platform for storing the spatial data and attributes of each model feature (e.g. surveying points, hotel and gas stations), while the ABM was developed using SI, specifically the Ant Colony Optimization (ACO), for modeling the behavior of the pavement evaluation crews in obtaining the minimum travelled distance for the data collection process.

GIS Platform

In this study, for performing the network analysis to obtain an optimized route, the use of Geographical Information Systems (GIS) is used. According to Gohari et al. 2012, for route planning systems, GIS provides the ability to manage

data spatially in layers and then perform network analysis for the purpose of route planning systems.

A GIS platform was developed to provide the spatial information of the variables to be considered as part of PDCRO: Hotel, Gas Stations and Milepost locations. For the road network and state boundaries, TIGER Lines ShapeFiles were downloaded from the New Mexico Resource Geographic Information System Program (RGIS). To properly analyze the distance traveled by an agent, all layers were projected to the Universal Transverse Mercator Zone 13 (UTM 13N) with the Geodetic Datum set as NAD 83.

ABM Approach

An Agent Based Model (ABM) using swarm intelligence assists in modeling the expected behavior of the agents throughout the GIS platform network. First developed by Marco Dorigo in 1992, ACO is a swarm intelligence approach which is inspired by the behavior of real ants (Dorigo 1992). The underlying theory of ACO is *"the foraging behavior of natural ant colonies which optimize their path from an origin to a destination by taking advantage of knowledge acquired by ants that previously traversed the possible paths, and the pheromone trail that these ants left behind as they traversed the paths to optimal solution."(Christodoulou 2010)*

Ant colony optimization fits better for optimization problems of discrete nature (Dorigo and Stutzle 2004) such as the "Traveling Salesman" problem and the "Backpack problem". The PDCRO is to be considered as one of discrete nature since the pavement evaluation crews will be choosing from a specific set of options to create their optimized path. In order to find a route from node i to node j artificial ants use a probabilistic function as following (Dorigo and Stutzle 2004):

$$p_{ij}^k = \begin{cases} \dfrac{\tau_{ij}^\alpha}{\sum_{l \in N_i^k} \tau_{il}^\alpha} & \text{if } j \in N_i^k; \\ 0 & \text{otherwise} \end{cases}$$

Where p_{ij}^k is the probability that the ant k chooses node j as its destination while there are other potential destinations is its neighborhood N_i^k, τ_{ij} is the amount of pheromone accumulated on the path from i to j and α is the pheromone magnification factor.

Upon completion of a colony's exploration the next one uses the pheromone information to start their search. In order to prohibit a pre-mature convegence the pheromone on the paths are reduced at specific intervals. The entire process continues until one or more stopping criteria are met. In this study, obtaining the minimum distance travelled was used as the stopping criteria.

Case Study

The proposed methodology was tested using information from the pavement data collection process performed by the University of New Mexico (UNM) for the New Mexico Department of Transportation (NMDOT) in summer 2012. UNM evaluates approximately 5,000 miles of pavement in New Mexico's route system. The evaluation includes a manual, visual evaluation of pavement surface distresses.

NMDOT uses the results of these evaluations to determine the serviceability of individual roads and make decisions on future maintenance and construction projects.

There were a total of 6 crews of pavement evaluators (two members each) designated to 6 different zones in New Mexico. Each of the six crews functioned as a mini Project Management team. They were to determine their route, and schedule it a week in advance. This included making decisions regarding lodging based on current location and the point at which they would start the data collection process the next day. The data obtained from UNM's data collection process was integrated in this study to assist with the GIS modeling and the SI based modeling. The data includes GPS coordinates, a survey deployed to pavement evaluation crews with regards to their decision making process and productivity at the data collection sites, GIS spatial data and attributes, among others.

Problem definition

The data collection sites in the state of New Mexico to be surveyed were limited to those located on I-40 and I-25, which intersect in Albuquerque. Both interstates where divided in two segments at the point of intersection, their intersections being the dividing point for each segment. Table 1 describes each segment, such as its length and total of milepost.

Table 1: Road segments consider for the PDCRO

Route	Location from ABQ	Segment ID	Length (km)*	Total mileposts
I-40	West	S-1	256 (159)	160
I-40	East	S-2	343 (213)	215
I-25	South	S-3	66 (41)	42
I-25	North	S-4	375 (233)	235
	Total	4	1040 (646)	652

* Numbers in parenthesis represents the values in miles

In order to demonstrate the methodology, in this study the objective function was to minimize the total distance travelled by the pavement evaluators. The initial point of departure of the agents was the City of Albuquerque. Initially, the pavement evaluation crew with two members selects a segment to begin the data collection process and then the crew proceeds by stopping at each milepost within the segment for evaluations. Finally, the crew selects hotels and gas stations as needed. When the data collection process of a segment is completed, the crew searches for the next segment to evaluate. It was assumed that the agent will travel in straight path between points.

The model considers that pavement evaluation crew would travel their way back to their starting point on each segment, by assuming they would travel each segment twice. In order to account for this, in the ACO it was assumed that the evaluation crew was to travel all the segments twice except for the first segment. Based on the number of miles traveled per day and the capacity of the fuel tank of the vehicle, each segment was zoned to a number of gas station zones and hotel zones. At each gas station zone an ant, which represents the pavement evaluation crew, selects

a gas station and at each hotel zone the pavement evaluation crew decides on its choice of hotel.

After the selection part, the evaluation crew calculates the distance from all the mileposts to the associated selected stop (whether it is a gas station or a hotel) and sees which milepost is the closest. In addition they have to decide if they have to go back to the same milepost upon return or can proceed to the next milepost from the stop. This decision is based on the distance from the stop to the closest milepost, to the next milepost and the distance between the two mileposts.

When the process is completed, which includes data collection, hotel selection and selection of gas stations, the evaluation crew represented by ants reports the total distance traveled based on these selections. When the entire population of ants completed their travel, the minimum length traveled is selected as the best solution. The ant, which has selected the best routes, is allowed to discharge pheromone on all the stops it has selected. The next population of ants starts to explore the solution space based on the updated pheromone initiation. This process continues until a stopping criterion is met which is reaching a specific number of generations in this study. The ACO algorithm used by authors to solve the PDCRO is shown in Figure 1.

```
Procedure
    While the stopping criteria is met
        n ← 1
        While n<population size
                SegOrder(n) ← construct the order of segments
                s ← 1
                While s<Number of Segments
                        g ← 1
                        While g<Number of gas zones in segment s (NG(s))
                                GasStation^n(s, g) ← select a gas station in gas zone g of segment s
                                End while
                        h ← 1
                        While h<Number of hotel zones in segment s (NH(s))
                                Hotel^n(s, h) ← select a hotel in gas zone h of segment s
                        End while
                end

                L(n) ← calculate the total length to travel
        End while
        L^{best} ← min( L(n))
        Update pheromone
    End while
End procedure
```

Figure 1: PDCRO Algorithm

RESULTS

The aforementioned algorithm was coded using Matlab. To run the algorithm, the total number of generations was set to 100 and at each generation 100 ants traveled the solution space to find the optimal results. After running the model, for approximately 7 minutes, the optimized travel length was obtained. The results indicate that the optimum solution for the PDCRO is to travel the segments (See

Table 1) in the following order: S-4 S-3, S-1, and S-2 for a total travel distance of 1834 km (1139.5 miles).
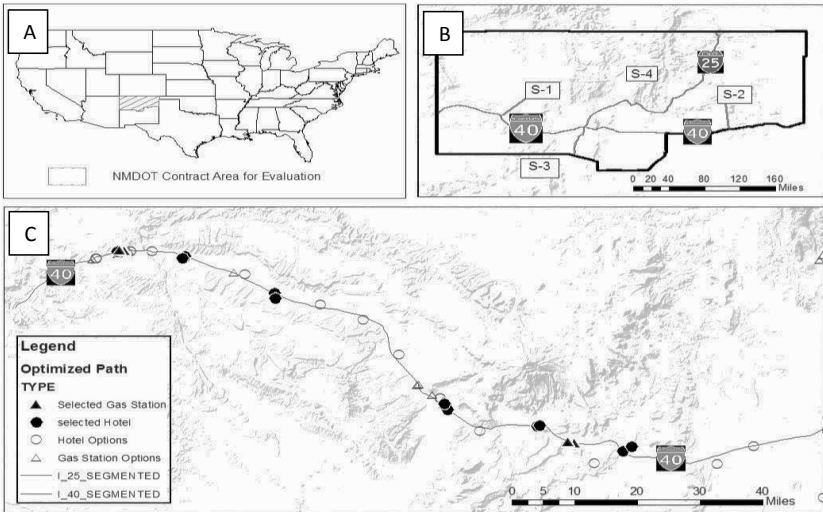
Figure 2: Destinations Selected by the PDCRO in Segment 1 (S-1)

Figure 2-c shows a view of the PDCRO results for Segment 1. As this figure shows, the pavement evaluation crew had a discrete amount of options of hotel and gas stations in their travel path. It can be observed that the selected hotels and gas stations by the pavement evaluators' crews were the closest alongside the interstate and to the point at which the crews collected their last data point at a given day. Once the crew evaluated all the mileposts within the segment, they had to return to the intersection point to start another segment. For this reason, it can be observed that there are hotels and gas stations selected close to each other.

Unfortunately, global optimal results were not available to verify the outcomes of the model since similar problem is not solved using a linear or un-linear programing probably due to run-time error and worst case complexity. However, in this study the authors tried to verify the solution improvement within the problem. Firstly, as described earlier, it is assumed that all the segments are travelled twice except for the first segment. Accordingly, the optimum results should show the longest segment as the one to begin with and the resulted optimum solution conforms to this fact.

In another attempt to verify the capability of the model, one of the stops was set on one of the mileposts to see whether the algorithm will choose this point to stop since it is definitely an optimal point. The other control point was to set a gas station and a hotel at the same stop to mandate another optimal stop. For both control points the best ant selected the estimated optimal solution which means that 'good' results are guaranteed.

CONCLUSION

Some agencies have incorporated the use of GPS and GIS software to properly maintain a database for their PMS. PDCRO presents a method for managing routing logistic and resources. The results of this research provide guidance to transportation agencies that could result in a more efficient data collection process.

A SI approach is proposed to solve the PDCRO. The model was able to produce 'good' results after searching only 4% of the solution space. The model was able to perform satisfactorily under control conditions, which shows there is a practical potential for the model. Future work will include parameter tuning.

As previously stated, research within this subject is very limited and it is recommended that it continues to grow. In this study, the objective function was to minimize the total distance travelled by the pavement evaluating crews. However, in the future, the authors suggest considering other parameters such as road geometry, signal settings, traffic data and cost which could be relevant for determining the optimum route selection. Also, it can be helpful to study how the ACO approach performs compared with traditional optimization methods, and expanding the pavement evaluation network in order to compare the results to historical data.

REFERENCES

Babak Farhang, M., Rubén, R., and Seyed Jafar, S. "Vehicle routing problem with uncertain demands: An advanced particle swarm algorithm." *Computers & Industrial Engineering*, 62, 306-317.

Bennet, C. R. (2008). "Data Collection Technologies For Pavement Management Systems." *7th International Conference on Managing Pavement Assets*, Transportation Research Board- National Academy of Sciences, Calgary, Alberta, 9.

Bennett, C. R. (2008). "Procuring pavement management data-collection services." *ROAD & TRANSPORT RESEARCH*, 17(3), 13-22.

Bogus, S., Migliaccio, G., and Cordova, A. (2010). "Assessment of Data Quality for Evaluations of Manual Pavement Distress." *Transportation Research Record*, 2170(1), 1.

Christodoulou, S. (2010). "Scheduling Resource-Constrained Projects with Ant Colony Optimization Artificial Agents." *Journal of Computing in Civil Engineering*, 24(1), 45-55.

Dorigo, M. (1992). "Optimization Learning and Natural Algorithms, PhD. Thesis." Politecnico di Miliano, Italia.

Dorigo, M., and Stutzle, T. (2004). "Ant Colony Optimization." MIT Press, Cambridge, Mass.

Feng, C.-M., Hsieh, C.-H., and Peng, S.-C. (2010). "Optimization of Urban Bus Routes based on Principles of Sustainable Transportation." *Journal of the Eastern Asia Society for Transportation Studies*, 8(0), 1137-1149.

Findley, D. J., Cunningham, C. M., and Hummer, J. E. (2011). "Comparison of mobile and manual data collection for roadway components." *Transportation Research Part C-Emerging Technologies*, 19(3), 521-540.

Flintsch, G. W., McGhee, K. K., National Research Council (U.S.). Transportation Research Board., National Cooperative Highway Research Program., American Association of State Highway and Transportation Officials., and United States. Federal Highway Administration. (2009). *Quality management of pavement condition data collection*, Transportation Research Board, Washington, D.C.

Gohari, A., H. A, A., M. G, H., S, K., and L, K. (2012). "Towards the Design of GIS-Based Routing System." *International Journal of Geoinformatics*, 8(2), 63-69.

Haughton, M. A. (2002). "MEASURING AND MANAGING THE LEARNING REQUIREMENTS OF ROUTE REOPTIMIZATION ON DELIVERY VEHICLE DRIVERS." *Journal of Business Logistics*, 23(2), 45-66.

Ibraheem, A. T., and Al-Razzaq Falih, D. a. A. (2012). "Applying Geographic Information System (GIS) for Maintenance Strategy Selection." *Engineering*, 4(1), 44-54.

Kong, R. S., Feng, J., Gao, R., and Zhou, H. B. (2012). "A New Route Optimization Scheme for Network Mobility: Combining ORC Protocol with RRH and Using Quota Mechanism." *Journal of Communications and Networks*, 14(1), 91-103.

Kumar, S. N., and Panneerselvam, R. (2012). "A Survey on the Vehicle Routing Problem and Its Variants." *Intelligent Information Management*, 4(3), 66-74.

Ladanyi, R. (2011). "OPTIMISATION OF SELECTIVE WASTE COLLECTION ROUTES ON THE BASIS OF GEOGRAPHICAL INFORMATION SYSTEM (GIS)." *Annals of the Faculty of Engineering Hunedoara - International Journal of Engineering*, 9(2), 87-90.

Lim, E. L. (2010). "Optimizing Waste Collection Routes." *Journal of Solid Waste Technology & Management*, 36(3), 234-243.

McGhee, K. H., National Cooperative Highway Research Program., American Association of State Highway and Transportation Officials., and National Research Council (U.S.). Transportation Research Board. (2004). *Automated pavement distress collection techniques*, Transportation Research Board, Washington, D.C.

Morian, D., Stoffels, S., and Frith, D. J. "Quality management of pavement performance data." *Proc., Pavement Evaluation Conference*, Transportation Research Board, 14.

Mosa, A. A., Abdalla, A. H., and Saeed, R. A. (2012). "Evaluation of MANEMO route optimization schemes." *Journal of Network & Computer Applications*, 35(5), 1454-1472.

Pajunas, A., Matto, E. J., Trick, M., and Zuluaga, L. F. (2007). "Optimizing Highway Transportation at the United States Postal Service." *Interfaces*, 37(6), 515-525.

Sahoo, S., Kim, S., Byung-In, K., Kraas, B., and Popov Jr, A. (2005). "Routing Optimization for Waste Management." *Interfaces*, 35(1), 24-36.

Wenjing, X., Dong, W., and Linbing, W. (2012). "A Review and Perspective about Pavement Monitoring." *International Journal of Pavement Research & Technology*, 5(5), 295-302.

Video-Based Highway Asset Recognition and 3D Localization

Vahid Balali[1], Mani Golparvar-Fard[2] and Jesus M. de la Garza[3]

[1]PhD Student, Vecellio Construction Engineering and Management Group, Via Department of Civil and Environmental Engineering, and Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA; PH (540) 235-6474; FAX (540) 231-7532; email: vbalali@vt.edu
[2]Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL; PH (217) 300-5226; FAX (540) 231-7532; email: mgolpar@illinois.edu
[3]Vecellio Professor, Vecellio Construction Engineering and Management Group, Via Department of Civil and Environmental Engineering, and Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA; PH (540) 231-7255; FAX (540) 231-7532; email: chema@vt.edu

ABSTRACT

A key element towards development of asset management program is an efficient data collection on high-quantity low-cost highway assets. Despite the importance, current practices of asset data collection are still manual and time-consuming. There is a need for a well-managed asset data collection that can provide useable asset inventories to departments of transportations (DOTs) for further analysis and condition assessment purposes. In this paper we present a novel video-based recognition and 3D reconstruction algorithm. Our method takes an input of video streams and combines 2D recognition with 3D reconstruction algorithms. Using a new Support Vector Machine (SVM) classifier and based on the color channels at pixel level, a set of bounding boxes are initially extracted. Using a Haar-based shape recognition algorithm, the 2D candidates are further categorized based on their shape. These candidates are placed into a texture+color recognition algorithm. The benefits and limitations of the method in detection, classification, and localization of multiple types of assets are discussed in detail.

INTRODUCTION

The significant expansion in size and complexity of highway networks in addition to the difficulties in data collection has made the National Academy of Engineering (NAE 2010) to identify the process of efficiently creating records of the locations and up-to-date status of the civil infrastructure as one of the grand Engineering Challenges of the 21st century. The American Society of Civil Engineers (ASCE 2011) has estimated that $2.2 trillion is needed to be spent over the next five years to repair and retrofit the U.S. infrastructure to a good condition. Despite the

significance, there is a lack of a reliable and up-to-date databases which can integrate geospatial, economic, and maintenance asset data. Such centralized databases can help DOTs to better prioritize different highway sections for maintenance and replacement planning purposes (Cheok et al. 2010). The significant size of the data which needs to be collected and the subjectivity and experience of the raters have an undoubted influence on the final assessments (Bianchini et al. 2010). Over the past few years, cheap and high-resolution video cameras and extensive data storage capacities have enabled DOTs to capture videos from highways and their assets on a truly massive scale. As a result, several new vision-based algorithms are developed that use these videos to automatically detect and classify highway assets in a discrete fashion. Despite the significant visual information that is embedded in these video frames, to date, the full application of these video frames for automated and simultaneous data collection of multiple types of assets is still untapped and unexploited by researchers. The intra-class variability of the assets and the static and dynamic occlusions can significantly challenge development of automated computer vision-based methods. One particular challenge is automated 3D reconstruction and recognition for different types of assets from these images. Figure 1 shows examples of two major categories of high-quantity, low-cost highway assets.

In this paper, we will address how our proposed computer vision based framework can detect, classify and spatially localize high-quantity low-cost highway assets. We will also discuss how the proposed computer vision based framework enables the transformative potential of using mounted cameras on an inspection vehicle as sensors and reporters of location and up-to-date status of assets and conditions, and can minimize the deficiencies of the current practices of infrastructure asset condition assessment.

RESEARCH BACKGROUND

Data collection, data management, and data integration are essential parts of a successful highway asset management program. The level of detail and accuracy of data collection primarily depend on the intended use of the data, yet in almost all cases can be classified into three categories (Flintsch and Bryant 2009): 1) Location of the asset; 2) Physical attributes: e.g., description of the asset, size, length; and 3) Condition: qualitative and generic: e.g., good or bad; quantitative and detailed: e.g., asset condition index. Most of the highway inventory data collection methods use one or more visual sensing methods to capture road inventory information. GPS, Inertia Measurement Unit (IMU), and Distance Measurement Indicator (DMI) are often used to provide accurate positional data for these visual sensing systems (Jalayer et al. 2012).
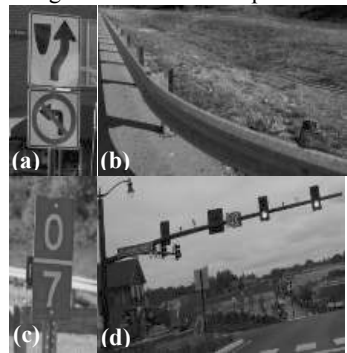
Figure 1: (a) and (c) assets that can be detected in 2D; (b) and (d) assets that cannot be detected from single image and need to be detected in 3D

The computer vision community has largely turned towards the recognition of object classes, rather than specific highway assets such as traffic signs. Both False Positive (FP) and False Negative (FN) rates have to be very low in highway asset management. That is why currently much of this work is still carried out by human operators. Detection of traffic signs as classified in Manual on Uniform Traffic Control Devices (MUTCD) is an area that has received considerable attention over the past few years. (Hu and Tsai 2011) presented a sign recognition algorithm that primarily focuses on speed limit signs. Their algorithm benefits from a probabilistic color model, the likelihood of sign locations in the images along with the traditional sign features. Others focus on features such as color, shape, geometrical and physical features along with text. More research is still ongoing on creating methods that can detect other types of assets beyond traffic signs such as mile posts, traffic signals, and guardrails.

In the AEC/FM community, (Golparvar-Fard et al. 2012a) proposed a new dense 3D reconstruction algorithm based on multi-view stereo and voxel coloring/labeling techniques which significantly improve the quality of the reconstructed models. (Brilakis et al. 2011) also proposed a structured videogrammetry for 3D reconstruction of existing highway bridges. The low resolution of Structure from Motion (SfM) point clouds may not be suitable for asset detection, localization, and condition asset. (Timofte et al. 2011) propose a new approach for 2D recognition and 3D localization of traffic signs. Despite the great performance reported, the recent algorithms mainly result in sparse 3D point cloud models and as a result may not be useful for 3D detection and classification of guardrails and light poles. In the context of infrastructure projects, (Golparvar-Fard et al. 2012b) proposed a new asset detection and recognition algorithm based on Semantic Texton Forest that can simultaneously segment an image and categorize assets.

Over all, there is a need for improvement on the state-of-the-art vision-based techniques for 3D reconstruction, detection and localization of assets. Particularly, a new environment needs to be created where the geometrical (3D) and appearance (2D image) information of the assets is integrated, providing a platform for development of joint recognition and localization of assets. In the following, our new method for creating such an integrated environment, plus joint 3D reconstruction and segmentation of assets from 3D point clouds is presented in detail.

PROPOSED HYPOTHESIZED FRAMEWORK

Using the video streams collected from the cameras mounted on the vehicle, a set of *thresholded* frames is initially identified. Each thresholded frame contains a set of candidates for low-cost high volume assets along the right side of the highway. Using a new SVM classifier and based on the color channels at pixel level, a set of bounding boxes are initially extracted wherein each bounding box potentially includes an asset. This stage passes all assets that are partially occluded, damaged or their signage is faded. Next, using a new shape recognition algorithm based on Haar-like features, the 2D candidates are further refined and categorized based on their shape. These candidates are placed into a new texture and color recognition algorithm, wherein by using a multiple binary SVM classifier, they are further classified into particular predefined types of assets. This step is mainly detecting assets which are recognizable in 2D. In

the meantime, using an improved image-based 3D reconstruction pipeline consisting of SfM and Multi View Stereo (MVS) algorithms, a point cloud model of the highway and all assets along is reconstructed and the images are geo-registered in a common 3D environment.

Color-Based 2D Candidate Detection

Given the speed of moving vehicles and the number of frames collected by the high speed camera, it is important to continuously assess at least the possibility of existing assets that are partially occluded, damaged or their signage is faded. The 2D asset problem is traditionally solved by one of the following approaches: 1) the selective extraction of windows of interest, followed by their classification; 2) exhaustive sliding window based classification. In the approach (1) small number of interest regions are selected in the images, through fast and cheap methods. These interest regions are then subjected to a more sophisticated classification. Such approach

Figure 2: Initial 2D Asset Candidate Extraction

risks overlooking 2D assets. Approach (2) considers all windows in the image. As the number of candidate windows is huge, classification process easily becomes intractable (Timofte et al. 2011). In the first step, as shown in Figure 2, by using a new SVM classifier and based on the color channels at pixel level, a set of pixels are initially selected as candidate locations of assets. Next, using the Maximally Stable Extremal Region (MSER) feature detection technique (Matas et al. 2004), the pixels are grouped in several regions as connected 2D asset candidates. By using a novel secondary thresholding SVM-based classifier, the regions are scaled to form enhanced bounding boxes. This stage is useful in the case where the asset is not well locally separable from the background. Ultimately a set of initial bounding boxes are extracted wherein each bounding box potentially includes an asset. While this algorithm returns very few FN (non-detected assets), it is designed to return a high number of FP (potential candidates for assets).

Shape-Based 2D Candidate Detection

2D assets have specific characteristic shapes. Hence, in this step, the initial 2D asset candidates are refined by using a shape recognition algorithm which filters out the FP of the previous color-based detector based on Haar-like features and ultimately categories assets into categories of similar 2D shape appearances (e.g., circle, diamond). In the proposed algorithm (Fig. 3), by first using a sliding window detector, the selected regions are convolved with multiple Haar-like filters (See Fig. 4), the pixel
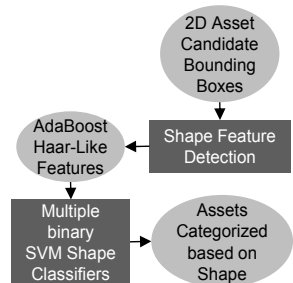
Figure 3: 2D Shape Recognition

intensities of adjacent rectangular regions are summed up and the differences are calculated to form the Haar-like features. These features are calculated very fast and are independent of different image resolutions and are robust to noise and changes in illumination. Next, the shape candidate is learned by selection of weak classifiers based

Figure 4: Haar Filters

on the joint Haar-like features using AdaBoost learning algorithm. Finally, the shape candidates are placed into a multiple one-against-all binary SVM shape classifier and various asset shapes are detected.
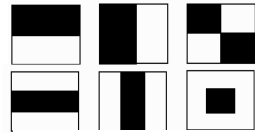
Color+Texture Based 2D Candidate Detection

2D Candidate clustered based on shape are placed into a new color and texture recognition algorithm. In this step, first using (Leung and Malik 2001) comprehensive radial and linear filter bank, a set of training candidate frames are convolved and the outcomes are clustered using K-means clustering approach. The formed cluster centers will be called Textons. Next, a histogram of these Texton centers per training frame will be created and using a dimensional reduction technique, the number of training samples is reduced. The same process is conducted on the color channels of the frames, wherein the color channels are initially converted into HSV to increase the robustness of the detection to changes in environment brightness and the same processes of clustering and histogram formation are performed. These histograms are combined and placed into a new multiple one-against-all texture+color binary SVM classifier where in the frames are further categorized into refined asset categories (e.g., circle-shape Speed Limit sign of 30 mph). One a novel frame goes through the pipeline of texture and HSV color detection, the histograms will be formed based on trained Textons and HSV cluster centers and ultimately the classifier will categorize assets into the most appropriate 2D asset category (Fig. 5).

Image-Based 3D Reconstruction

The collected video frames are also used to generate a dense 3D point cloud model that can be used for: 1) identification of 3D assets; 2) localization of 2D and 3D assets; and finally 3) generating an augmented reality environment for data management and contextual awareness applications, plus automated condition assessment analyses. The first steps in image-based reconstruction as observed in Fig. 6 are to perform real-time feature detection and matching on the video frame. In the proposed approach, a GPU implementation of SIFT feature detection (Lowe 2004) is created to track and match features in all the frames. The camera tracker is initialized with the relative pose of three views, given feature correspondences in them. The correspondences
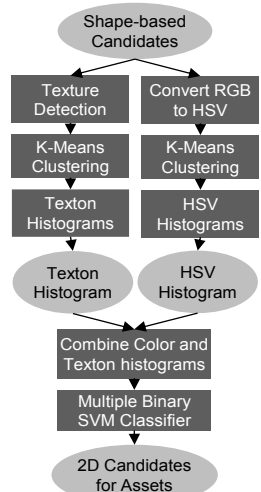
Figure 5: Robust Texture and Color Recognition

among feature points will be triangulated using the computed camera poses. Additional poses will be computed with RANSAC and the constraints generated in the previous stage to map 2D feature points to 3D correspondences. Next, new world points will be re-triangulated using new views as they become available. In order to avoid accumulated drifts, the model is periodically reinitialized with a new set of three views and GPS information is used for ego-motion calculations of the inspection vehicle. Finally the 3D correspondences are transformed into 3D world coordinate system. The outcome of this step is a sparse 3D point cloud along with camera parameters transformed into the 3D coordinate system. Next, the dense reconstruction and registration pipeline proposed in (Golparvar-Fard et al. 2012a) which consists of MVS and Voxel Coloring/Labeling steps is validated by feeding the outcome of the improved SfM algorithm and generating dense point cloud model.
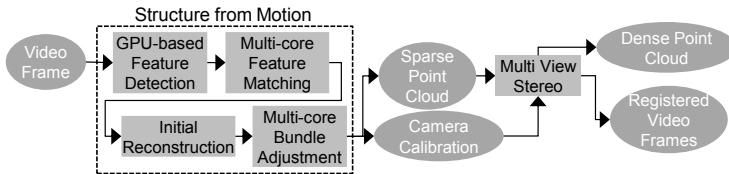
Figure 6: Image-based 3D reconstruction (inspired by (Golparvar-Fard et al. 2010))

VALIDATION METRICS

The validation process of the proposed approach for *color-based 2D candidate detection* consists of implementing the thresholded frame extraction method and then selecting the potential frames that contains assets. All true positive (TP), true negative (TN), false positive (FP), and false negatives (FN) detections will then be counted manually, and used to construct precision-recall graphs to represent the detection performance. The results are evaluated based on a benefit/cost analysis, where benefit is defined as the reduction of frames sent to the next process, and cost as the speed of processing and the number of assets missed over all assets. As it shown in Figure 7, extraction of connected components from a threshold image is used for highway asset detection. It generates a lot of FPs while the number of FNs are low. The thresholded image is obtained from a color image.

Similar to the evaluation in previous step, for *shape-based 2D candidate detection* the results will be evaluated and are used to further define the number of Haar filters required, the impact of the scale, and ultimately the kernel types of the binary SVM classifiers.

The validation process for *color and texture* approach consists of
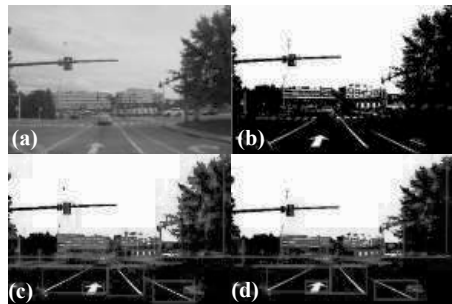
Figure 7: Color-based extraction: (a) Colorful image; (b) Binary image; (c) Binary image with bounding boxes; (d) Binary image with detected bounding boxes.

counting all TP, TN, FP and FN detections, and forming confusion matrix for one-against-all classification algorithm. The classification results can help determine the required number of filters in the filter bank and Textons and HSV color cluster centers to result in highest accuracy determined through confusion matrix.

The accuracy and sensitivity of *3D reconstruction* will be validated in two steps. A set of video frames that contain assets and are representative of the whole dataset with a 95 percent confidence will be considered for reconstruction. The dimensions of these assets are predetermined and through experiment, the deviation between the reconstructed and actual measurements will be used for assessment of recall (completeness of reconstruction) and precision (accuracy of reconstruction) with 80% confidence in the results. The results of 3D point cloud are shown in Figure 8. The outcome is shown in 4 dimensional augmented reality (D4AR) visualization models.
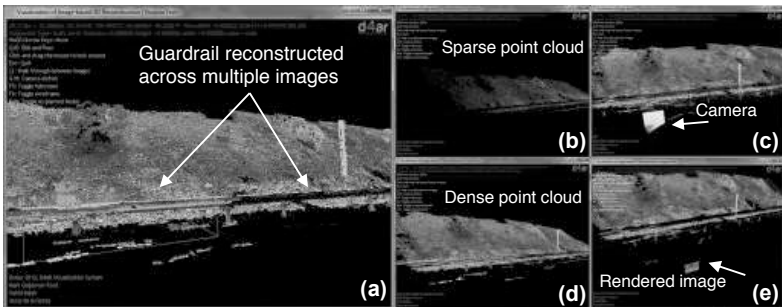
Figure 8: 3D image-based reconstruction: (a) reconstructed 3D point cloud model highlighting the guardrail; (b) sparse point cloud model; (c) camera viewpoint; (d) dense point cloud model; (e) rendered image over same camera frustum shown in (d)

CONCLUSION

The proposed methods as well as the overall pipeline lend themselves to applications in different contexts and with different imagery. For instance, the optimization for picking the best thresholds has a general formulation and parameters can be adapted based on training data. The number of annotations, the ratio between FPs and FNs, and the precision of segmentation can be used to set parameters. Following a principle of spending little time on the bulk of the data, and keeping a more refined analysis for the promising parts of the images, the proposed system combines efficiency with good performance. The integer linear optimization formulation for selecting the optimal candidate extraction methods and the standard sliding window approach are found to be complementary to the proposed detection based on fast extracted candidates. The detected assets and their types are visualized in an augmented reality environment which enables remote walk-throughs for condition assessment purposes. The validation of several parts of this application are still ongoing. It can also improve public access to information about location and condition of road assets. It can also foster information sharing and exchange among different agencies as well as DOTs.

ACKNOWLEDGEMENTS

REFERENCES

ASCE, A. S. o. C. E. (2011). "National Infrastructure Report Card."

Bianchini, A., Bandini, P., and Smith, D. W. (2010). "Interrater Reliability of Manual Pavement Distress Evaluations." *Journal of Transportation Engineering*, 136(2), 165-172.

Brilakis, I., Fathi, H., and Rashidi, A. (2011). "Progressive 3D reconstruction of infrastructure with videogrammetry." *Automation in Construction*, 20(7), 884-895.

Cheok, G., Franaszek, M., Katz, I., Lytle, A., Saidi, K., and Scott, N. (2010). "Assessing Technology Gaps for the Federal Highway Administration Digital Highway Measurement Program.", Construction Metrology and Automation Group, Building and Fire Research Laboratory, National Institute of Standards and Technology (NIST).

Flintsch, G. W., and Bryant, J. W. (2009). "Asset Management Data Collection for Supporting Decision Process." FHWA, U.S. Department of Transportation, Federal Highway Administration, FHWA-IF-08-018.

Golparvar-Fard, M., Balali, V., and de la Garza, J. M. (2012b). "Segmentation and Recognition of Highway Assets Using Image-Based 3D Point Clouds and Semantic Texton Forests." *Journal of Computing in Civil Engineering*.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2010). "D4AR – 4 Dimensional augmented reality - tools for automated remote progress tracking and support of decision-enabling tasks in the AEC/FM industry." *Proc., the 6th Int. Conf. on Innovations in AEC.*

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012a). "Automated operation-level tracking of progress using unordered daily construction photographs and IFC as-planned models." *Journal of Computing in Civil Engineering(ASCE), In Press.*

Hu, Z., and Tsai, Y. (2011). "Generalized Image Recognition Algorithm for Sign Inventory." *Journal of Computing in Civil Engineering*, 25(2), 149-158.

Jalayer, M., Gong, J., Zhou, H., and Gordon, C. "Mobile Terrestrial Laser Scanning for Highway Inventory Data Collection." *Proc., ASCE International Workshop on Computing in Civil Engineering*, 545-552.

Leung, T., and Malik, J. (2001). "Representing and Recognizing the Visual Appearance of Materials using Three-dimensional Textons." *Int. J. Comput. Vision*, 43(1), 29-44.

Lowe, D. G. (2004). "Distinctive Image Features from Scale-Invariant Keypoints." *Int. J. Comput. Vision*, 60(2), 91-110.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). "Robust wide-baseline stereo from maximally stable extremal regions." *Image and Vision Computing*, 22(10), 761-767.

NAE, N. A. o. E. (2010). "Grand Challenges for Engineering. NAE of the National Academies.", NAE, National Academy of Engineers.

Timofte, R., Zimmermann, K., and Luc Van, G. (2011). "Multi-view traffic sign detection, recognition, and 3D localisation." *Journal of Machine Vision and Applications (MVA), Springer-Verlag*.

Network-Wide Assessment of Transportation Systems Using an Epidemic Spreading Methodology

Seyed Hossein Hosseini Nourzad, S.M.ASCE[1], Anu Pradhan, A.M.ASCE[2]

[1]PhD Candidate, Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA 19104: PH (215) 429-0700; FAX (215) 895-1363; email: hnourzad@drexel.edu
[2]Assistant Professor, Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA 19104; PH (215) 571-3540; FAX (215) 895-1363; email: pradhan@drexel.edu

ABSTRACT

Network-wide assessment of transportation systems is a crucial task in traffic planning and management. Several simulation models (including macroscopic, mesoscopic, microscopic, as well as hybrid modeling) and theoretical models (e.g., macroscopic fundamental diagram) provide such assessments with different levels of details. Although such models present good insight into dynamics of traffic flow (e.g., the relationship between speed and density) they do not discover the interrelationship between traffic flow dynamics and topology of the network. In this paper, the authors have attempted to macroscopically simulate congestion propagation on a regional road network based on the concept of epidemic spreading models (e.g., susceptible-infected-susceptible method). The results showed that at a critical threshold a transition occurs from a free flow phase to a congested phase. In addition, the simulation results verified that such a phase transition happens at the proposed mathematical threshold based a topological characteristic (i.e., largest eigenvalue of adjacency matrix). However, the differences in the nature of congestion propagation and epidemic spreading suggest that further studies are needed to develop a hybrid method (i.e., macro-meso-micro) which mixes the proposed method with existing traffic simulation methods.

INTRODUCTION

Transportation networks are an important class of civil infrastructure system. The incapacity or destruction of such networks could severely impact social and economic security of our societies. On the other hand, efficient transportation networks provide reliable flows of essential products and services which facilitate social connectivity and impact the nation's economy. Therefore, it is important to design and maintain efficient transportation networks. A substantial amount of research studies have been performed to model different types of transportation, such as road networks (Chan et al. 2011), railway network (Doménech 2009), and air traveling network (Guimerà et al. 2005). In particular, road network is the common mode of transportation around the world (Papacostas and Prevedouros 2001; Chan et al. 2011; Chiu et al. 2012).

At present, different types of methods based on various levels of fidelity (i.e., micro, meso, and macro) are being used for traffic simulation on a given road network (Florian 2008; Balakrishna et al. 2012; Chiu et al. 2012). Among them, microscopic simulation methods provide more accurate results at the expense of detailed data and higher computational costs. Mesoscopic simulation methods are less demanding in terms of computational costs compared to microscopic methods, but still require detailed data, such as demand between different pairs of zones and road topology/geometry. Macroscopic simulation methods use aggregate data to predict speed of vehicles based on vehicle density. However, none of the existing methods is able to directly model congestion propagation. To simulate congestion propagation using the existing methods, one needs to increase the demand loading on the network, and evaluate the emergence and propagation of congestion within the network.

In this manuscript, the authors proposed a method to model the congestion propagation in a given road network inspired from epidemic spreading in a social network, which is a network of individuals linked through some interactions. Such a proposed method requires a basic knowledge of the network structure (e.g., connectivity), average rate of congestion transfer from one node to its adjacent nodes, and average rate of congestion relief in a node. The proposed method can predict the phase transition behavior of congestion on a road network. The results were compared with the phase transition of epidemic spreading in a social network. Finally, future research directions were proposed to be development of a hybrid method (i.e., macro-meso-micro), which mixes the proposed method with existing traffic simulation methods and validating the results employing real-life road networks.

BACKGROUND RESEARCH

In general, congestion propagation is an important problem in the road transportation domain (Papacostas and Prevedouros 2001). To analyze such a problem, it is necessary to understand the network conditions at different levels of congestion. Recently, macroscopic fundamental diagram (MFD) has been developed to depict the macroscopic relationship between average speed and density for urban networks (Daganzo and Geroliminis 2008). In addition, some researchers modeled vehicle movement on the road network using either analytical or simulation based methods (Lam and Xu 1999). Analytical methods contain mathematically proven properties and simple driver behaviors for small networks. However, in the case of large real-world networks, the analytical methods cannot represent detailed complexity of the network and driver behavior (Ziliaskopoulos et al. 2004). Therefore, simulation-based methods were developed, such as DynaMIT (Balakrishna et al. 2008), DYNASMART (Mahmassani 2001), VISTA (Ziliaskopoulos et al. 2004), DynusT (Chiu et al. 2012), Dynameq (Florian 2008), AIMSUN (Barceló and Casas 2006), TransModeler (Balakrishna et al. 2012), INTEGRATION (Van Aerde et al. 1996), METROPOLIS (De Palma and Marchal 2002). The simulation-based methods require detailed information about the demand and structure of the road networks.

Based on level of details, current simulation-based methods can be categorized into three classes: 1) macroscopic, 2) mesoscopic, and 3) microscopic.

The macroscopic methods utilize physical concepts such as fluids theory without considering individual vehicle (Lighthill and Whitham 1955). The microscopic methods model individual vehicle entities, decisions and interactions at time step, which is as small as one-tenth of a second. The microscopic models require detailed calibration of model parameters, and have high computational resources (Smith et al. 1995). However, the mesoscopic methods combine elements from both microscopic and macroscopic methods and each time step can be as large as sixty seconds (Mahmassani 2001; Chiu et al. 2012).

While all of the existing methods (e.g., macro, meso, or micro) are developed to simulate traffic on a road network, they are unable to directly simulate congestion propagation within the network. Currently, to simulate congestion propagation, one has to increase the demand step by step, and repeat the simulation until the congestion emerges and propagates within the network. Moreover, the existing theoretical methods (e.g., MFD) focus more on the traffic flow dynamics rather than the network topology. Therefore, network-wide impacts of the topological characteristics (i.e., eigenvalues of adjacency matrix) on the traffic flow dynamics (i.e., congestion) require more studies.

RESEARCH APPROACH

This section describes the research approach. In general, the research question is whether the congestion propagation within a road network is analogous to an epidemic spreading in a social network or not, at least from a large-scale perspective. The proposed approach is based on the method of epidemic spreading in a social network. The aim of epidemic spreading modeling is to reproduce the actual dynamics of infections, and to understand impacts of network topology on the epidemic spreading (Pastor-Satorras and Vespignani 2001; Boccaletti et al. 2006). In this paper, we modeled the behavior of a social network using the **S**usceptible–**I**nfected–**S**usceptible (SIS) model in which infections do not confer immunity to their survivors (e.g., tuberculosis and gonorrhea). In a social network, the individuals and their interactions with other individuals are modeled as network nodes and links. At a given time step, each node can be in only one of the two states: a) susceptible (prone to the infection), or b) infected (gets infected and spreads the infection). The probability of infection propagation from an infected node to its adjacent susceptible node is referred to as the spreading rate (equal to δ). In addition, once a node gets infected, the probability of recovering from the infection and returning to the susceptible state is referred to as the recovery rate (equal to μ) (Pastor-Satorras and Vespignani 2001).

In the SIS model, the infection starts from one or a few nodes within a network. Such an infection will then propagate from the infected nodes to the neighboring nodes with the rate of δ. At the same time, some of the infected nodes will recover with the rate of μ. At each time step, the next state of each node is a function of its current state and of the states of its neighbors on the network. Therefore, dynamics of the network depends on both the spreading and recovery rates and more specifically on their ratio (i.e., $\sigma = \delta / \mu$). In fact, there is an epidemic phase transition (i.e., a critical value of the ratio, $\sigma = \sigma_c$) (Wang et al. 2003). For $\sigma > \sigma_c$, the infection persists in the network, while for $\sigma < \sigma_c$ it does not. Wang et al. (2003)

proved that the epidemic threshold of any given network (σ_c) is equal to $1/\lambda_{1,A}$, in which $\lambda_{1,A}$ is the largest eigenvalue of adjacency matrix of the network.

The congestion propagation within a road network has some similarities to the epidemic spreading in a social network. For a given road network, the intersections and road segments are network nodes and links. When a node is congested, the congestion may spillback to upstream nodes with the average propagation rate (δ). At the same time, some of the congested nodes may become decongested with the average relief rate (μ). In other words, each node can be in only one of the two states: a) uncongested (referred to as susceptible), or b) congested (referred to as infected). Therefore, network-wide impacts of congestion propagation can be simulated using an SIS model. The quantity of interest is the critical threshold (i.e., σ_c) in a road network associated to the phase transition from free-flow to congested phase. For mathematical details of the SIS models, interested readers can refer to (Pastor-Satorras and Vespignani 2001; Boccaletti et al. 2006).

RESULTS

This section presents the results of the two case studies: 1) the epidemic spreading, and 2) the congestion propagation. The differences between the two case studies and the effectiveness of the proposed approach are described next.

4.1 Epidemic spreading: To verify the mathematical solution of the epidemic threshold, several simulations were performed using package described in (Muchnik et al. 2007). For the generated social network, 1) degree distribution was in the form of $P(k) \sim k^{-1.5}$, 2) number of nodes was 10000, 3) recovery rate was assumed to be constant (μ=0.5) for all the simulations, 4) spreading rate changed ranging from δ=0.0105 to δ=0.02, and 5) largest eigenvalue of the adjacency matrix ($\lambda_{1,A}$) was equal to 90.90. Therefore, the epidemic threshold was supposed to be $\sigma_c = 1/\lambda_{1,A} = 0.011$. The time steps of simulation were assumed to be 1 day.

Figure 1 shows the number of nodes that were infected in each time step in the three different simulations. For $\sigma < 0.011$ (i.e., Figure 1 (a)), the number of infected nodes became equal to zero, which means that the epidemic died after a period of time. For $\sigma \approx \sigma_c$ (Figure 1 (b)), there was a fluctuation in the number of infected nodes, which means that the network was in the transition phase. In Figure 1 (b), the average number of infected nodes was not zero, because the ratio σ is slightly greater than σ_c. For $\sigma > 0.011$ (i.e., Figure 1 (c)), the number of infected nodes reached a constant nonzero number, which means that the epidemic persisted in the network. Therefore, the simulation results illustrate that there is a phase transition behavior in epidemic spreading within a social network. The knowledge of such behavior could help epidemiologists assess the spreading of different epidemics and mitigate the risks of such epidemics.

(a) δ=0.00525, σ = 0.0105

(b) δ=0.0055, σ = 0.011

(c) δ=0.01, σ = 0.02
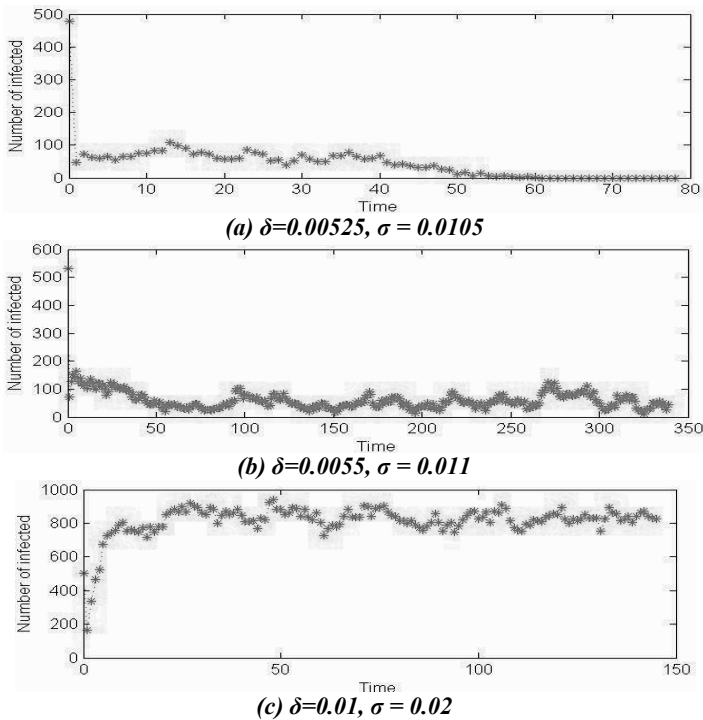
Figure 1. Simulation results for epidemic spreading

        4.2 Congestion propagation: The congestion propagation on a given regional road network from Island of Guam was simulated using the proposed method. The Guam network was selected for this study because it is an Island with no through traffic, which means that the origins and destinations of all the travels are within the network. To run the simulation, the proposed method was developed in MATLAB using the package described in (Muchnik et al. 2007). For the Guam road network: 1) the number of nodes was equal to 539, and 2) the number of links was equal to 1183. The largest eigenvalue of the adjacency matrix ($\lambda_{1,A}$) was equal to 2.819. Therefore, the mathematical solution for the transition threshold (σ_c) was supposed to be equal to 0.355. The time steps of simulation were assumed to be 5 minutes. The simulation was performed for a period of 4 hours (i.e., 240 minutes or 48 time steps).

        Figure 2 shows number of nodes that were congested at each time step for three different simulations. For σ < 0.355 (i.e., Figure 2 (a)), the number of congested nodes was zero, which means that the congestion disappeared after a period of time. For σ ≈ σ_c (Figure 2 (b)), there was a fluctuation in the number of congested nodes, which means that the network was in the transition phase. In contrast to Figure 1 (b), in Figure 2 (b) the average number of congested nodes was zero, because the ratio σ is slightly smaller than σ_c. For σ > 0.355 (i.e., Figure 2 (c)), the number of congested nodes reached a constant nonzero number, which means that the congestion persisted

in the network. Hence, the results show that there is a phase transition behavior in congestion propagation within a road network. Finally, a sudden drop can be seen in the number of congested nodes between time step 0 and 1 because the ratio of propagation rate to the relief rate is less than one. This means that many of the congested nodes are relieved due to high value of relief rate.

(a) δ=0.044, σ = 0.177

(b) δ=0.089, σ = 0.355

(c) δ=0.177, σ = 0.709
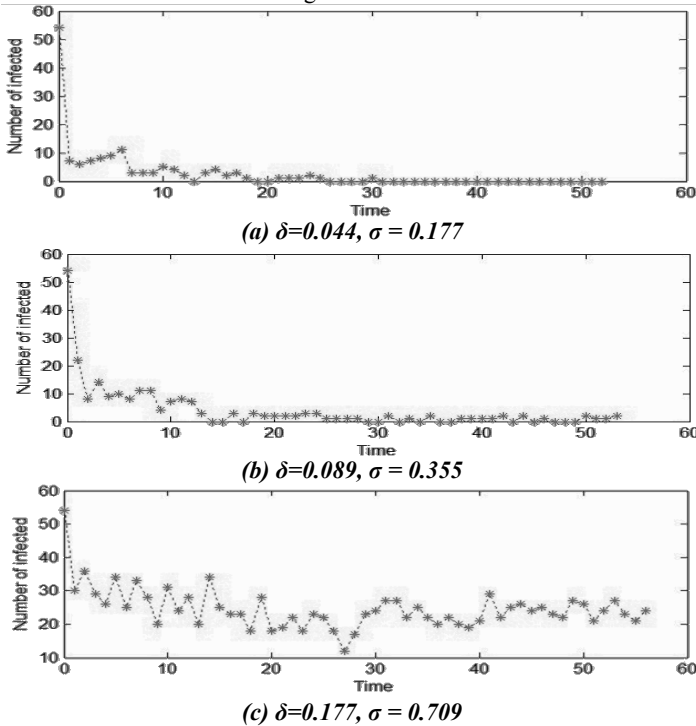
Figure 2. Simulation results for congestion propagation on Guam road network

4.3 Discussions: There are some differences between road and social networks. For example, links in road networks (i.e., roads) are physical links with currents (i.e., flows) running on top of the links. If the roads cannot handle the incoming traffic, traffic congestion on a road network will be generated by vehicles that move on the roads. The congestion may then spillback to the upstream of the road, leading to the congestion spreading within the network. Therefore, the rate of congestion propagation (δ) for each link depends on the traffic flow at each time step. The traffic flows, depend on the distribution of the travel demand. Since travel demand is heterogeneously distributed within a road network, the rate of congestion propagation also changes for different links and time steps.

In addition, the congestion relief of a road (μ) depends on the state of its neighboring node. For example, a congested road cannot become decongested as long as its downstream roads are congested. In contrary, the recovery of a node in a social network is independent of the state of its adjacent nodes. For example, an individual

may recover from an infection while its adjacent individuals are still infected. Therefore, the recovery mathematical equation (for epidemic spreading) has to be adjusted for congestion propagation to consider the influences of the adjacent nodes.

CONCLUSIONS

A network-wide method was proposed to simulate congestion propagation on a road network macroscopically. The goal of this research was to predict network-wide transition from free-flow to congested phase based on topological characteristics of the road network. The proposed method was built on the epidemic spreading concept using simulation methods. The two case studies were performed: 1) the epidemic spreading on a synthetic social network with 10000 nodes, and 2) the congestion propagation on the road network from Island of Guam. For the both of case studies, simulation results showed that the phase transition was occurred at the critical threshold, $\sigma_c = 1/\lambda_{1,A}$. However, the differences between nature of congestion propagation and epidemic spreading demonstrate the need for a more detailed simulation of vehicle movement. Further research studies can be performed to present a hybrid method (i.e., macro-meso-micro). Such a hybrid method mixes the capability of the proposed method in simulating the congestion propagation with the capability of the existing simulation methods in simulating traffic flow. In future works, the hybrid method will be employed for different real-life cases to validate the proposed method. Moreover, estimation of congestion propagation rate (δ) and relief rate (μ) is an open question for future research.

ACKNOWLEDGMENT

REFERENCES

Balakrishna, R., Morgan, D., Yang, Q. and Slavin, H. (2012). Comparison of Simulation-Based DTA Approaches for Planning and Operations Management.

Balakrishna, R., Wen, Y., Ben-Akiva, M. and Antoniou, C. (2008). Simulation-based framework for transportation network management in emergencies, Transportation Research Record: 2041(-1): 80-88.

Barceló, J. and Casas, J. (2006). Stochastic heuristic DA based on AIMSUN microscopic traffic simulator, Transportation Research Record 1964(-1): 70-80.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. and Hwang, D. U. (2006). Complex networks: Structure and dynamics, Physics Reports 424(4–5): 175-308.

Chan, S. H. Y., Donner, R. V. and Lämmer, S. (2011). Urban road networks - spatial networks with universal geometric features? The European Physical Journal B 84(4): 563-577.

Chiu, Y.-C., Nava, E. J. and Hu, H.-H. (2012). A Temporal Domain Decomposition Algorithmic Scheme for Efficient Mega-Scale Dynamic Traffic Assignment - An Experience with Southern California Associations of Government DTA, SCAG.

Daganzo, C. F., & Geroliminis, N. (2008). An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological, 42(9), 771–781.

De Palma, A. and Marchal, F. (2002). Real cases applications of the fully dynamic METROPOLIS tool-box: an advocacy for large-scale mesoscopic transportation systems, Networks and Spatial Economics 2(4): 347-369.

Doménech, A. (2009). A topological phase transition between small-worlds and fractal scaling in urban railway transportation networks? Physica A: Statistical Mechanics and its Applications 388(21): 4658-4668.

Executive.Order (1995). Critical Infrastructure Protection, Federal Register: 61.

Florian, M. (2008). Models and software for urban and regional transportation planning: the contributions of the center for research on transportation, INFOR: Information Systems and Operational Research 46(1): 29-50.

Guimerà, R., Mossa, S., Turtschi, A. and Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles, National Academy of Sciences 102(22): 7794-7799.

Lam, W. H. K. and Xu, G. (1999). A traffic flow simulator for network reliability assessment, Journal of advanced transportation 33(2): 159-182.

Lighthill, M. and Whitham, G. (1955). On kinematic waves. I. Flood movement in long rivers, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 229(1178): 281-316.

Mahmassani, H. S. (2001). Dynamic network traffic assignment and simulation methodology for advanced system management applications, Networks and Spatial Economics 1(3): 267-292.

Muchnik, L., Itzhack, R., Solomon, S. and Louzoun, Y. (2007). Self-emergence of knowledge trees: Extraction of the Wikipedia hierarchies, Physical Review E 76(1): 016106.

Papacostas, C. S. and Prevedouros, P. D. (2001). Transportation engineering and planning, Prentice Hall.

Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks, Physical Review Letters 86(14): 3200-3203.

Smith, M., Duncan, G. and Druitt, S. (1995). PARAMICS: microscopic traffic simulation for congestion management, Dynamic Control of Strategic Inter-Urban Road Networks, IEE Colloquium on, IET.

Van Aerde, M., Hellinga, B., Baker, M. and Rakha, H. (1996). INTEGRATION: An overview of traffic simulation features, Transportation Research Records.

Wang, Y., Chakrabarti, D., Wang, C. and Faloutsos, C. (2003). Epidemic spreading in real networks: An eigenvalue viewpoint. Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on, IEEE.

Ziliaskopoulos, A. K., Waller, S. T., Li, Y. and Byram, M. (2004). Large-scale dynamic traffic assignment: Implementation issues and computational analysis. Journal of Transportation Engineering 130(5): 585-593.

Data-driven Active Parking Management

Zhen (Sean) Qian[1] and Ram Rajagopal[2]

[1]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305; email: seanqian@stanford.edu

[2] Corresponding author, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305; email: ramr@stanford.edu

ABSTRACT

Parking is an essential component of urban transportation systems. The price and availability of parking could considerably influence travelers' choice of when to leave, where to park and which mode to choose. We propose dynamic pricing and sensing for parking in order to manage the traffic and reduce social costs. Cutting-edge sensing technology makes it possible to obtain novel parking information regarding real-time usage of parking facilities. The optimal prices are set with respect to the real-time parking occupancy. The effectiveness of pricing-sensing interaction is examined in several parking structures on the Stanford University campus.

INTRODUCTION

Downtown parking is a challenging issue for travelers, and it is also among the most common problems for transportation planners and operators. Efficient parking management can mitigate both traffic and parking congestion. It contributes to significant reduction in emissions and fuel consumptions, and increase in the social welfare. The objective of this research is to assess the social costs of automobile network caused by parking and to maximize the benefits of parking management through pricing and sensing.

Parking is an essential component of urban transportation systems. Statistics show that a typical vehicle is parked 23 hours each day, and parking cruising time represents up to 40% of the total travel time (Axhausen et. al. 1994). The price and availability of parking could considerably influence travelers' choice of when to leave, where to park and which mode to choose.

Both pricing and sensing have been widely applied in transportation management, and either pricing or sensing in parking is not a novel idea in the literature. However, their relationships are not fully understood yet and they were usually investigated in the steady-state context. In fact, there are quite a few descriptive and empirical studies on parking policies and integrated parking management (e.g., Thompson et. al. 1998, Vianna et. al. 2004). Although qualitative guideline for parking management or pricing is available, theoretical studies on parking modeling are few in the literature. As for parking modeling in dynamic context, Arnott et. al. (1991) are among the first to study the parking problem by dynamic user equilibrium. They embedded the parking problem in the well-known morning commute model (Vickrey 1969) to show that parking fee itself can be efficient in increasing social welfare. More recently, the parking capacity allocation, parking accessibility and parking fees are examined for a dynamic network with one roadway bottleneck and two parking areas (Qian et. al. 2011,2012). Those parking studies do not consider parking cruising time as part of the social costs.

Equipped with parking sensors, the technology of collecting approximate parking occupancy information has been available for dynamic parking management

systems in the last two decades. In some cities, travelers are provided with parking information in the form of variable message signs (VMS) or traffic radio. The information is usually regarding the number of available spaces in certain parking lots. However, in most cases, parking prices are unchanged from day to day, and are not provided to travelers as part of the parking information. Although occupancy information in VMS could be used to optimize the parking reliability (Mei et. al. 2012), several empirical studies and simulation show that the benefits of providing only parking availability information may be marginal in congested network conditions (e.g., Asakura and Kashiwadani 1994). Unfortunately, there is a lack of underlying theoretical parking models to provide real-time optimal pricing that fully uses valuable parking information and data.

This paper proposes a generic model for setting the optimal parking prices that minimized the total system cost. The time-varying prices can be set solely upon the real-time parking occupancy data. Numerical examples are provided using traffic data collected from the campus of Stanford University to illustrate the relationship between the optimal dynamic parking pricing and the traffic patterns.

PRICING AND SENSING

Pricing has been an efficient instrument to adjust travel demand. However, most of the current parking prices are not time-varying, and may not be adaptive to real-time traffic demand. On the other hand, the cutting-edge sensing technology makes it possible to obtain unprecedented parking information, including real-time occupancy of all lots, even with precise location of available spots. By providing this information to the travelers, unnecessary parking cruising for fully occupied lots can be reduced. However, this improvement could be marginal to the transportation system, if travelers are only aware of the parking usage information. This is because such information cannot predict the future usage precisely nor take actions on managing the traffic, and thus information provision alone is not efficient to reduce parking congestion, especially for those most congested lots. Pricing can be used to control the traffic efficiently, along with the information provision.

Here we propose a pricing-sensing loop control system for a set of parking lots in the transportation network. The idea is shown in Figure 1. At any time stage, the parking system takes measurement of the real-time parking usage. The parking data are then analyzed to forecast the future usage and demand, and determine the optimal real-time parking prices for each lot. Furthermore, the parking usage (existing and predictive) information and parking prices are sent to travelers through smart phones, websites, variable message signs, radios and so on. With real-time information, travelers make their decisions on which lot to choose and when to leave. In the next time stage, the up-to-date parking usage is detected and reported to the parking system to update how users response to the prices and information, which is considered to set the new optimal prices. Overall, parking pricing and sensing jointly serve as a dynamic stabilized controller of the traffic.

THE MODEL

Suppose there are n parking lots for commuters to choose from. They are located on their way to the office sequentially. Each lot represents a parking structure

(namely parking garage) or a block of on-street parking spots. Such a sequence of parking lots are defined as $\{n, n-1, ..., 2, 1\}$ and illustrated in Figure 2.

Decision: choice of time and/or lots

Sensing: information
Management: forecast, prices
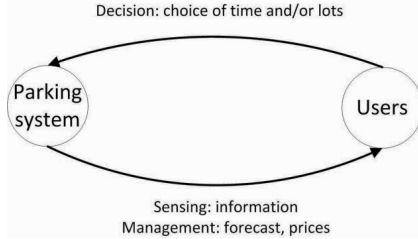
Figure 1. Interaction between pricing and sensing.

Figure 2: A sequence of parking lots

There are in all N travelers who leave home for the office. Each traveler that arrives in the system chooses a parking lot i and then walks to the office. We use $\lambda(t)$ to denote the traveler arrival rate at time t. Therefore, $\int_0^T \lambda(t)dt = N$. where T is the commuting ending time. Let τ_i^d denote the driving time of a traveler from his home to lot i, τ_i^w the walking time from lot i to the office. The *composite* travel time $\tau_i = \tau_i^d + \tau_i^w$ of a traveler is defined as the sum of the time from his home to lot i and his walking time to the office. Due to the parking structure, $\tau_1 < \tau_2 < \cdots < \tau_{n-1} < \tau_n$ and $\tau_1^d > \tau_2^d > \cdots > \tau_{n-1}^d > \tau_n^d$. If no fees are imposed in the parking lots and users are provided real-time information that every parking lot has vacant spaces, then clearly lot 1 is the most desirable and lot n is the least desirable. Each parking lot i has a capacity of K_i spaces. A parking fee is charged to travelers upon their arrivals, represented by $p_i(t)$ at time t. In addition, $k_i(t)$ represents the cumulative number of arrivals to lot i by time t, namely the effective occupancy. $k_i(t) \leq K_i, \forall t$.

The generalized travel cost of travelers choosing parking lot i and departing home at time t, $C_i(t)$ (as $t + \tau_i^d$ is his arrival time to lot i), consists of the parking fee, composite travel time and parking cruising time,

$$C_i(t) = p_i\left(t + \tau_i^d\right) + \alpha\left(\tau_i + f_i\left(k_i(t + \tau_i^d)\right)\right) = \tilde{p}_i(t) + \alpha(\tau_i + f_i(\tilde{k}_i(t))) \qquad (1)$$

where α is the average value of time (VOT) for the traveler population. where $\tilde{p}_i(t)$ and $\tilde{k}_i(t)$ represents the parking price and the effective occupancy of lot i for a traveler departing home at time t (since the driving time τ_i^d is fixed). $f_i(t)$ denotes the expected vehicle cruising time at parking lot i for a traveler arriving at time t. This cruising time is in fact closely related to the parking occupancy and the type of parking information provided to the users. Consider the case where travelers are provided with the number of vacant spaces in each parking lot with, however, no exact spot location information (this is the most common case in the current parking management system). The expected parking cruising time $f_i(\tilde{k}_i(t))$ reads,

$$f_i(\tilde{k}_i(t)) = \frac{\varepsilon_i}{1 - \frac{\tilde{k}_i(t)}{K_i}} \qquad (2)$$

where ε_i the average time spent on searching one parking space in lot i. This special form of cruising time is consistent with Axhausen et. al. (1994) and Anderson and Palma (2004). A generic cruising time function was discussed in Qian et. al (2012).

The time-dependent travel demand, i.e. $\lambda(t)$ over t, is assumed to be predetermined, or as can be detected by traffic count sensors. Travelers choose their respective parking spaces to minimize their travel costs, and after day-to-day experience, eventually the traveler population achieves a User Equilibrium flow pattern defined as follows.

Definition of User Equilibrium (UE) without departure time choices
Given the location, capacity and parking fee of all parking lots, travelers are aware of the real-time occupancies of each lot, and thus the expected cruising time after a sufficient amount of commuting experience (or by estimated cruising time provided to travelers), and the eventual flow patterns are such that 1) their generalized travel costs are the same after their day-to-day adjustments (or by assuming fully rational choices); 2) No traveler can unilaterally change his parking choice to reduce his generalized travel cost.

Under UE behavior assumption, different parking prices lead to different UE flow patterns, and thus different network performance, represented by the total travel cost (including the cruising time, driving time and traffic congestion). We are most concerned about the System Optimal (SO) pricing schemes that can lead to the minimal total cost. The resultant UE flow pattern under the SO prices is referred to as the SO flow pattern. To solve the System Optimum (SO) parking pricing, we first seek the optimal parking flow pattern that minimizes the total cost. The optimization problem reads,

$$\min_{\lambda} TC = \alpha \sum_{i=1}^{m} \left(\int_0^{\tilde{k}_i(T)} f_i(x)dx + \tilde{k}_i(T)\tau_i \right) \tag{4}$$

$$s.t. \sum_{i=1}^{m} \lambda_i(t) = \lambda(t), \forall 0 \leq t \leq T \tag{5}$$

$$\tilde{k}_i(T) = \int_0^T \lambda_i(t)dt, \forall 1 \leq i \leq m, i \in \mathbb{N} \tag{6}$$

$$\lambda_i(t) \geq 0, \forall 1 \leq i \leq m, i \in \mathbb{N}, \forall 0 \leq t \leq T \tag{7}$$

where λ is the vector of parking flow pattern, $\{\lambda_1(t), \lambda_2(t), ..., \lambda_m(t)\}, \forall t$. m is the number of lots being used when the commute ends. T is the commute ending time. Therefore, $\tilde{k}_i(T)$ is the number of spaces finally occupied in lot i at the departure time T. Equation 5 states that the flow feasibility, i.e. the sum of the parking flow for all lots is equal to the total commuting flow at any departure time t. Equation 6 computes the terminal occupancy of each lot by integrating its own flow pattern.

Clearly, the optimal pattern λ^* is not unique. This is because if the terminal occupancies $\{k_1^*(T), k_2^*(T), ..., k_m^*(T)\}$ are SO solutions, then any λ satisfying the feasibility constraints (i.e. Equations 5, 6 and 7) would be an optimal pattern. It is easy to show that the SO solution of $\{k_1^*(T), k_2^*(T), ..., k_m^*(T)\}$ is unique if the cruising time functions $f_i(k), \forall i$ are strictly monotone. The optimality condition is (MTC is short for *marginal travel cost*),

$$\tau_i + f_i(k_i^*(T)) = \frac{MTC}{\alpha}, , \forall 1 \leq i \leq m, i \in \mathbb{N} \tag{8}$$

$$\sum_{i=1}^{m} k_i^*(T) = N \tag{9}$$

Equation 8 shows that, any additional traveler using lot i, regardless of his departure time, can increase the total generalized travel time by his composite travel time τ_i and

the marginal parking cruising time $f_i(k_i(T))$. For any departure time, System Optimum requires the MTC of any traveler is equalized.

The non-uniqueness of the SO parking flow pattern indeed offers much flexibility to set different dynamic pricing schemes for a variety of management goals, while all those pricing schemes could be SO. In fact, we can obtain any desired arrival rates to parking lots by setting appropriate SO pricing schemes. To see this, suppose the desired flow pattern is $\bar{\lambda} = \{\bar{\lambda}_i(t), \forall i, t\}$ that satisfies the feasibility condition 5-7. Let $\bar{k}_i(t) = \int_0^t \bar{\lambda}_i(t')dt', \forall i, t$, which are desired time-varying occupancies with respect to the departure time t. We use $p_i^*(t)$ to represent the SO pricing schemes that realizes the flow pattern $\bar{\lambda}$.

Without loss of generality, we can always set the parking price of one of the lots being used as the *base* price for any departure time t, and that lot is referred as the base lot, B(t). Note that the base lot could change over time, as any lot may not necessarily be used over the entire commuting time under the targeted flow pattern. In order to obtain a unique optimal price for a desired flow pattern, we will need to add two more reasonable constrains. First, we assume the parking price at any lot does not decrease over time. Second, for any time t, the parking price of a lot not being used is higher than the price that makes it get a share, by a certain level, say σ. A sufficient great σ is necessary to direct travelers to use other parking lots in reality. For any lot i being used at time t, its parking price is

$$p_i^*(t) = p_j^*(t) + \alpha(\tau_j + f_j(\bar{k}_j(t))) - \alpha(\tau_i + f_i(\bar{k}_i(t))), j = B(t), \quad (10)$$

If lot i is not used by any traveler at time t, its price is set to,

$$p_i^*(t) = p_j^*(t) + \alpha(\tau_j + f_j(\bar{k}_j(t))) - \alpha(\tau_i + f_i(\bar{k}_i(t))) + \sigma, j = B(t), \quad (11)$$

The procedure to solve the optimal pricing is described as follows.

Initialization; Solve $\{k_i^*(T)\}$; Determine $\{\bar{k}_i(t)\}$ and $\{\bar{\lambda}_i(t)\}$; Determine initial price p_0 and price increment σ. Determine $S(t)$: the set of parking lots being used at time t.

Boundary condition; Determine the base lot for the first time step, and set its price p_0. Use Equations 10 and 11 to determine the initial parking price for other lots.

For every time step t

Determine the base lot B(t), and $p_{B(t)}^*(t) = p_{B(t)}^*(t-1)$; let j=B(t);

Use Equation 10 to calculate $p_i^*(t)$ for $i \in S(t), i \notin B(t)$

Use Equation 11 to calculate $p_i^*(t)$ for $i \notin S(t), i \notin B(t)$. If $p_i^*(t) < p_i^*(t-1)$,

set $p_i^*(t) = p_i^*(t-1)$;

If $B(t) \neq B(t-1)$, then $p_i^*(t) = p_i^*(t) - min_l\{p_l^*(t) - p_l^*(t-1)\}$_for any lot i.

Output $\{p_i^*(t)\}, \forall i, \forall t$.

PARKING SENSOR SYSTEMS

We here use micro-radar sensors to detect the presence of the vehicles for every parking spot. The micro-radar sensor incorporates an extremely low power, wide-band radar. The compact sensor transmits high frequency RF pulses, and the RF pulses are bounced off a target object, and the return pulses are measured by a time-gated RF mixer. RF reflections are analyzed to produce presence, distance, and motion measurements. Micro-radar sensors are capable of detecting and distinguishing objects in motion from objects that are stationary, as well as large

objects from small objects, and thus can accurately detect parking vehicles. The sensor is battery powered, which can last for 10 years.

The parking sensors essentially obtain the real-time parking availability information. They talk directly to base stations, which further send real-time parking data to centralized servers. Using aforementioned parking models, we can set the SO parking prices solely based on real-time parking occupancies. The optimal prices balance the demand to each lot to achieve the system optimum. For any lot with an occupancy higher than or equal to the optimal terminal occupancy, an extremely high parking fee should be imposed.

NUMERICAL EXPERIMENTS

We choose a real parking network on the campus of Stanford University. There are four major parking garages located on the (only) major arterial road from the western entrance of the campus. Almost all of the travel demand heads for the central part of the campus. The properties of those parking lots are listed in Table 1. The driving time and walking time is estimated approximately based on the distance measured in Google Maps. Currently, all lots are available for commuters with certain parking permits. There are a few spaces (less than 5%) deployed with meters that are specially used for visitors. The parking fees here are set to be consistent with the permit price on the daily basis. The VOT α is set to 35 US dollars per hour (approximately the average hourly pay rate on campus) and $\beta = \alpha/4$. We also collected 15-min traffic count data upstream of the inbound major arterials. The time horizon for this analysis is 7:00am-10:00am. Now we compare the UE solution with current prices and the SO solution. Their terminal occupancies are shown in Table 2. In fact, their difference in terms of final occupancy is fairly small. The SO reduces the total cost by 0.32%.

Table 1. The properties of the four parking lots.

	Lot S-2	Lot 21	Lot 17	Lot 18 & 20
Driving time τ_i^d (Seconds)	180	120	60	0
Walking time (Seconds)	0	300	600	900
Parking capacity (spots)	736	188	505	608
Parking fee for UE case ($)	6	6	2	2
Parameter ε (Seconds)	5	5	5	5
Parameter ε' (Seconds)	1800	1800	1800	1800

The results are not surprising. The SO pricing directs several travelers to use the closer lots compared to UE solution, since those travelers can save a bit of cruising time provided with occupancy information. When the travel demand is fixed over time (with very little flexibility), it is certainly possible that the optimal pricing and information provision do not make significant difference in free-flow travel time and cruising time. However, as we discussed before, such SO pricing scheme can effectively adjust the flow pattern in any desired rates, which is more advantageous than only working with the cruising time cost and free-flow travel time cost. An intensive parking flow rate in a short period may yield the same total cruising time cost as a mild flow rate with longer commuting period. However, the latter results in much less roadway congestion than the former.

Here we investigate the SO parking flow pattern by implementing the commonly used flow proportionality strategy. The flow patterns under both UE and

SO are shown in Figure 3. The flow proportionality strategy produces the occupancy that is proportional to the terminal occupancy at any departure time t. Figure 4 plots the SO parking prices against time and real-time occupancy, When the occupancy is light or mild, the parking fee for the four lots are $7, $5, $2.5 and $0 from the closest to the farthest. While the occupancy increases to a certain level (except the closest one that is set as the base price), the parking price increase quickly. The price of the farthest one grow fastest, because the flow proportionality strategy inherently favors the farther one by setting a low price for it and high price for the closer ones in the first place, so as to maintain the constant parking market share for any time. When the occupancies of all lots are medium to high (e.g. 0.7-0.85 in this case), the closer one is seriously under-used. Therefore, the price at the farther lots should be raised in order to direct travelers to the closer ones later on.

Table 2. Terminal occupancies under UE and SO.

	Lot S-2	Lot 21	Lot 17	Lot 18 & 20	Total cost ($)	Reduction
Capacity	736	188	505	608	-	-
UE	726	182	488	398	9,203	-
SO	731	186	495	384	9,174	0.32%
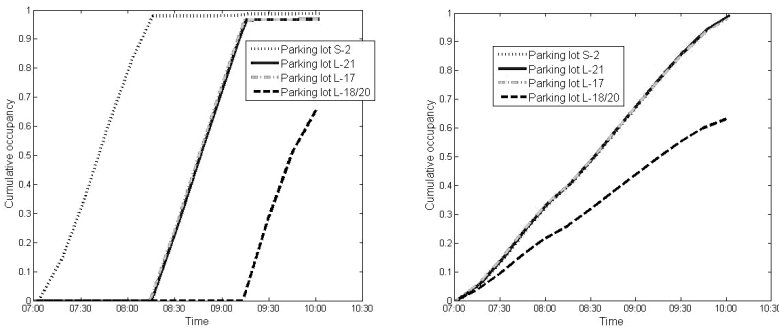
Figure 3. Parking flow patterns under UE and SO.
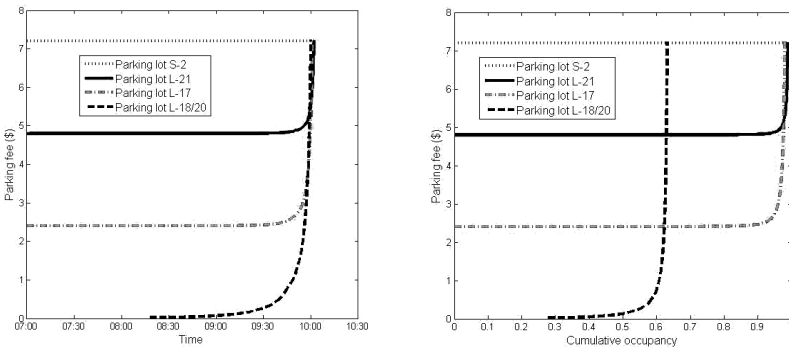
Figure 4. Optimal time-varying parking prices (flow proportionality).

CONCLUSIONS

This paper investigates dynamic parking pricing scheme and parking information provision to travelers for sequential parking lots. Suppose travelers are aware of the parking facilities, and they can obtain the real-time parking occupancy and pricing information for all the parking lots. They make parking location choices to minimize their generalized travel cost. We consider a general parking cruising time function as part of the travelers' generalized travel cost in addition to the driving time and walking time. We first derive the parking flow pattern, and then further solve the optimal parking prices with the corresponding optimal parking flow pattern.

It is found that the SO solutions to minimize the total cost (including cruising time and composite travel time) are not unique. The terminal occupancy under SO is usually close to that under UE where the parking prices are not significantly different between lots. However, the non-uniqueness of the SO parking flow pattern indeed offers much flexibility to set different dynamic pricing schemes for a variety of management goals, while all those pricing schemes remain SO. One important management strategy is to price the lots in a way that roadway congestion caused by parking cruising is minimized. In fact, data-driven SO pricing schemes are fully capable of achieving targeted flow patterns at each parking lot or area.

REFERENCES

Anderson, S. P. and A. de Palma. The economics of pricing parking. Journal of Urban Economics, vol. 55, no. 1, 2004, pp. 1–20.

Arnott, R., A. de Palma, and R. Lindsey. A temporal and spatial equilibrium analysis of commuter parking. Journal of Public Economics, vol. 45, 1991, 301–335.

Asakura, Y. and M. Kashiwadani. Effects of parking availability information on system performance: a simulation model approach. In Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994. aug-2 sep 1994, pp. 251 –254.

Axhausen, K., J. Polak, M. Boltze, and J. Puzicha. Effectiveness of the Parking Guidance Information System in Frankfurt/Main. Traffic Eninggering and Control, vol. 35, no. 5, 1994, pp. 304 – 309.

Mei, Z., Y. Tian, and D. Li. Analysis of Parking Reliability Guidance of Urban Parking Variable Mesage Sign System. Mathematical Problems in Engieering, vol. doi:10.1155/2012/128379, 2012, p. 10 pages.

Qian, Z. S., F. E. Xiao, and H. Zhang. The economics of parking provision for the morning commute. Transportation Research Part A: Policy and Practice, vol. 45, no. 9, 2011, pp. 861 – 879.

Qian, Z., F. Xiao, and H. M. Zhang. Managing Morning Commute with Parking. Transportation Research Part B, accepted, 2012.

Qian, S. and R. Rajagopal. Modeling parking pricing with occupancy information. working paper, 2012.

Thompson, R.G., K.Takada,and S. Kobayakawa. Understanding the demand for access information. Transportation Research Part C, vol. 6(4), 1998, 231–245.

Vianna, M. M. B., L. da Silva Portugal, and R. Balassiano. Intelligent transportation systems and parking management: implementation potential in a Brazilian city. Cities, vol. doi:10.1016/j.cities.2004.01.001, no. 2, 2004.

Vickrey, W. Congestion Theory and Transport Investment. American Economic Review, vol. 59, 1969, pp. 251–261.

The Impact of Predictive Cruise Control on Traffic Flow and Energy Consumption

Sehyun Tak[1] and Hwasoo Yeo[2]

[1]Smart Transportation System Laboratory, Department of Civil and environmental Engineering, KAIST,291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea PH (042) 350-3674; email: taksehyun@kaist.ac.kr
[2]Smart Transportation System Laboratory, Department of Civil and environmental Engineering, KAIST,291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea PH (042) 350-3634; email: hwasoo@kaist.edu

ABSTRACT

A vehicle Predictive Cruise Control system has been developed to improve the fuel efficiency of vehicle and traffic flow performance based on the asymmetric traffic theory. The Predictive Cruise Control system consists of four parts: (1) Deceleration based Safety Surrogate Measure, (2) Adaptive Cruise Control, (3) Multi-vehicle measurement, and (4) Predictive Cruise Control. Adaptive Cruise Control basically decides the acceleration/deceleration action based on the estimated deceleration-based safety surrogate measure of the first leader vehicle. Then, Predictive Cruise Control adjusts the acceleration/deceleration action based on the multi-vehicle measurements, which represent the future traffic condition of the subject vehicle. The developed system is tested by simulating the real vehicle trajectories from the NGSIM data and comparing the results with real following pattern. It was found that the newly proposed Predictive Cruise Control system can contribute to energy consumption and traffic flow performance, because it can effectively suppress the shockwave from the downstream and remove the unnecessary deceleration and acceleration action.

Introduction

It is known that many vehicle accidents are due to the carelessness, and human error. Therefore, many efforts have been made to improve safety by limiting a vehicle's maximum speed, ensuring the safe distance, and pre-alert system. Accident rate has been reduced because of these efforts. However, in recent years, accident reduction rate is at a standstill. And, an innovative system is needed to improve the human's safety in vehicle operations. The other issues related to the human driving behavior are traffic flow stability and energy consumption. Overreactions of drivers on the change in the leader vehicle propagate unnecessary shockwave and also make

negative impact on traffic flow and fuel consumption (Davis, 2004). So, reducing the impact of shockwave is important to improve the traffic flow stability.

One of the strong alternatives to solve these problems is Adaptive Cruise Control (ACC). Adaptive cruise control system enhances regular cruise control by automatically changing the velocity and spacing (Zhou and Peng, 2005). ACC vehicle is controlled based on the information, such as velocity of preceding vehicle and space between two vehicles. Such information is gathered from in-vehicle devices, communications with adjacent vehicles, and transportation infrastructure. In general, for ACC system, there are two types of spacing policies: constant space headway (or time headway) policy and variable space headway policy. Constant space headway policy determines the distance between preceding vehicle and following vehicle based on the assumption that two inter-vehicle distance and velocity have a linear relationship. This policy is almost entirely focusing on the velocity and space headway at a certain time. So, headway is linearly increased as the speed increases. Variable space headway policy considers not only the speed and spacing but also the situation faced by a certain vehicle, such as reduction of the relative velocity, increase in the relative velocity, and temporary shock wave. So, variable space headway policy-based vehicle has different space headway depending on the situation, even though space and velocity remains the same.

Out of the two types of space headway policies, the constant space headway policy (or time headway policy) is the one that is frequently used in ACC designs (Ioannou and Chien, 1993, Ioannou and Stefanovic, 2005, Darbha and Rajagopal, 1999). Constant space headway-based ACC vehicle has a capability to improve driver safety and increase the road capacity. However, Constant space headway policy has some disadvantages. First, constant space headway policy-based ACC vehicle uses more fuel, because it responses too sensitively to the changes in the preceding vehicle. Second, constant space headway policy-based ACC vehicle cannot mitigate the shock wave even though it can prevent the propagation of shockwave. So, to overcome these shortcomings, we develop the new Predictive Cruise Control (PCC) and evaluate the potential effects of the PCC on traffic stability and energy consumption.

Predictive Cruise Control Strategy

The newly proposed Predictive Cruise Control (PCC) consists of four parts: (1) Deceleration based safety surrogate measure, (2) Adaptive cruise control, (3) Multi-vehicle measure, and (4) Predictive Cruise Control as shown in Figure 1. Adaptive Cruise Control system provides the automatic tracking of the first leader vehicle. The optimal acceleration action and attenuation of preceding traffic is indicated as PCC based on the multi-vehicle measure, which indicates the situation of the preceding vehicles. The proposed PCC differs from ACC in that predicts the near future of the subject vehicle by utilizing the multi-vehicle measure. Detailed strategies are described as follows.

(1) Development of new safety surrogate measure for single vehicle.

The selection and design of the measure, such as spacing policy for determining the action for acceleration, deceleration, and constant speed, are considered as the prior factors in ACC (Swaroop and Rajagopal, 2001). ACC determines the collision risk of the subject vehicle based on the measure or spacing policy. In order to increase the efficiency and reliability of ACC system, various measures such as time headway, space headway, safety factor spacing, time to collision, and nonlinear spacing policy have been developed. However, the current measures for ACC have some limitations to reflect the characteristics of individual drivers and vehicle such as maximum braking performance, preferred acceleration rate, and etc. These limitations may cause uncomfortable environment for the driver due to the mechanical motion of the vehicle.
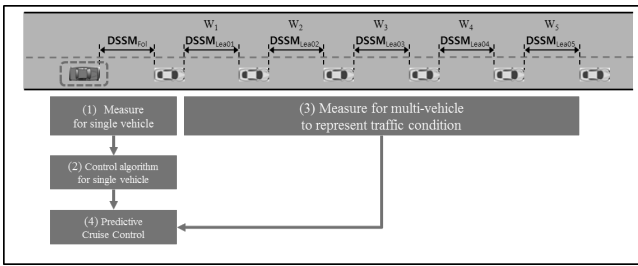
Figure 1. Overview of the components of the proposed Predictive Cruise Control

To enhance the current ACC, we propose a new measure, Deceleration-based Safety Surrogate Measure (DSSM), which is based on the safety condition (Yeo and Skabardonis, Yeo, 2008, Yeo et al., 2011). Traffic safety is one of the most important issues to motivate researches into ACC and collision warning or prevention technology is the essential function in ACC system. Thus, to meet the safety requirements, we develop the new measure based on the concept of safety surrogate measure, which is the widely used safety performance indicator. Deceleration-based Safety Surrogate Measure (DSSM) can be expressed as the function of the information obtained from V2V communication technology. To represent the collision risk based on the vehicle performance, the DSSM for ACC is defined as the function of braking performance as follows.

$$K= \left[x_n(t)-x_{n-1}(t)+s_{n-1}\right]+[2 \cdot v_n(t)+a_n(t) \cdot \tau] \cdot \frac{\tau}{2} - \left[v_{n-1}(t)/2+\frac{(a_{n-1}(t)+b_{max, n-1})}{4} \cdot \frac{(a_{n-1}(t)-b_{max, n-1})}{L_{n-1}}\right] \cdot \frac{(a_{n-1}(t)-b_{max, n-1})}{L_{n-1}} +$$
$$\left[v_n(t)/2+a_n(t) \cdot \tau+\frac{(a_n(t)+b_{max, n})}{4} \cdot \frac{(a_n(t)-b_{max, n})}{L_n}\right] \cdot \frac{(a_n(t)-b_{max, n})}{L_n} \tag{1}$$

$$b_n(t)=b_{max, \ n-1}(t) \cdot \frac{[v_n(t)+a_n(t) \cdot \tau]^2}{\left[2 \cdot K \cdot b_{max, \ n-1}(t)+v_{n-1}(t)^2\right]} \tag{2}$$

$$DSSM=\frac{b_n(t)}{b_{max.n}} \tag{3}$$

where $a_n(t)$ is the acceleration rate of following vehicle at time t, $a_{n-1}(t)$ is the acceleration rate of leader vehicle at time t, $b_{max.n-1}$ is the maximum braking rate of leader vehicle, which represents the vehicle's mechanical deceleration performance,

$b_n(t)$ is the needed deceleration rate of following vehicle to avoid the accident at time t, $b_{max.n}$ is the maximum braking rate of following vehicle, $v_{n-1}(t)$ is the speed of leader vehicle at time t, $v_n(t)$ is the speed of following vehicle at time t, L_{n-1} is the maximum variation of acceleration of leader vehicle, L_n is the maximum variation of acceleration of following vehicle, $v_n(t+\tau)$ is the expected speed of following vehicle after τ, $x_{n-1}(t)$ is the location of leader vehicle at time t, $x_n(t)$ is the location of following vehicle at time t, τ is the perception reaction time, and s_{n-1} is the length of leader vehicle.

ACC system always works with a human driver. Therefore, the first step in designing a vehicle control strategy for the application of ACC system is to develop the measure based on the driving behavior characteristics. The DSSM reflects the collision risk between two consecutive vehicles based on human and machine characteristics. The driver's comfort can be increased by using a preferred acceleration rate as the individual vehicle's $a_{n-1}(t)$ value, preferred braking rate as the individual vehicle's $b_{max.n}$ value, and preferred variation of acceleration rate as the individual vehicle's L_n value.

(2) Development of control algorithm for single vehicle.

The DSSM value represents the collision risk of the subject vehicle. The DSSM value larger than one represents that the subject vehicle cannot avoid the accident when leader reduce the speed with its maximum braking rate. The DSSM value greater than or equal to 1.2 means that the following vehicle cannot physically avoid collision when the leading vehicle stops with full deceleration. In reality, the drivers' patterns show that most of drivers conduct severe deceleration when DSSM value is greater than or equal to 1.2. DSSM value of 0.85 can be considered as the slightly dangerous situation and the driver shows the average number of approximately 0.85. By referring these values, the control algorithm is developed as shown in Figure 2.
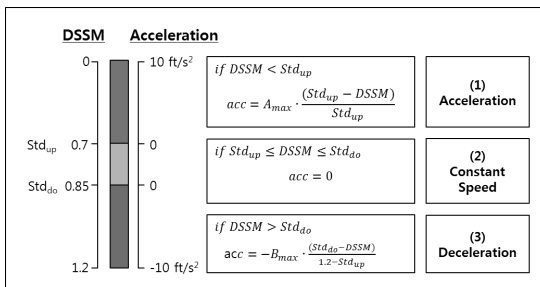
Figure 2. Longitudinal control algorithm of Adaptive Cruise Control (ACC) for single vehicle

Figure 2 shows the longitudinal control algorithm of ACC for single vehicle. The acceleration and deceleration rate is determined based on the calculated DSSM value. The longitudinal control algorithm consists of three parts. (1) The subject vehicle increases the speed proportional to the difference between Std_{up} and calculated DSSM. Std_{up} is the reference point for determining the acceleration action and 0.7 is

used for that value in this study. The frequent acceleration action is prevented and the intensity of acceleration rate is increased by decreasing the Std_{up}. (2) The Subject vehicle maintains the current speed when calculated DSSM is between Std_{up} and Std_{do}. (3) The subject vehicle reduces the speed proportional to the difference between Sdt_{do} and calculated DSSM. Std_{do} is the reference point for determining the deceleration action and 0.85 is used for that value in this study. By changing the Std_{do} value, the degree of risk taking is determined. For example, the driver who wants to reduce the collision risk can decrease the Std_{do}. On the other hand, the driver who does not wants frequent braking can increase the Std_{do} up to 1.0. Std_{up} and Std_{do} can be changed depending on the user preferences and system objectives.

(3) Development of algorithm on traffic condition.

To predict the future traffic condition, the proposed control algorithm uses the platoon concept and a multi-vehicle measure. The measure of the preceding vehicles can represent the future state of the subject vehicle. For example, the speed reduction of the foremost vehicle in a platoon means that the subject vehicle would reduce its speed in near future. In the situation based on the concept of DSSM, an increase in foremost DSSM, which is the $DSSM_{lea05}$ in Figure 1, has the possibility of increasing the subject vehicle's DSSM, which is the $DSSM_{fol}$ in Figure 1, even though the $DSSM_{lea01}$ is low. Conversely, low level of foremost DSSM leads the subject vehicle to low level of DSSM even though $DSSM_{lea01}$ is high. To indicate the future impact of the preceding vehicles on the subject vehicle, multi-vehicle DSSM is defined as:

$$DSSM_{Co} = \sum_{i=1}^{5} D_i \cdot W_i \qquad (4)$$

where $DSSM_{co}$ is the multi-vehicle DSSM, D_i is the DSSM of the i-th vehicle, and W_i is the weighting factor of the i-th vehicle.

D_i is also calculated by using equation (1), (2), and (3). W_i can be adjusted depending on the purpose of the user. If a user wants to response to D_1 more sensitively, which is the nearest future for the subject vehicle, the user can increase the W_1. If the user wants to response to the further future, the user can increase the W_5. In the next chapter, in order to evaluate the performance of the proposed PCC method we assigned 0.5 for W_1, 0.3 for W_2, 0.1 for W_3, 0.075 for W_4, and 0.025 for W_5.

(4) Development of Predictive cruise control

Figure 3 shows the PCC algorithm by using $DSSM_{co}$ and $DSSM_{fol}$. As shown in Figure 3, the control space is divided into 10 regions, where each region corresponds to the control mode of ACC for subject vehicle. The PCC system compares the measurements of $DSSM_{fol}$ and $DSSSM_{co}$ to the diagram to select the control mode. $DSSM_{fol}$ represents the DSSM value of the subject vehicle, and $DSSM_{co}$ represents the weighted average of DSSM for preceding vehicles. Region (1), (5), and (10) represent the no change mode of PCC. No change means that subject vehicle determines the deceleration/acceleration rate with the same level of single vehicle mode, so called ACC mode. Region (6), (8), and (9) means that DSSM of subject

vehicle will be reduced in the near future. So, subject vehicle accelerates more than the calculated acceleration rate in ACC mode by increasing the Sdt_{do} value. In region (2), (3), and (4), subject vehicle reduces the acceleration rate less than ACC mode by reducing the Std_{up}. In this region, subject vehicle increases the speed when only considering the nearest leader vehicle. However, the subject vehicle reduces the acceleration rate, because high DSSM of preceding vehicle may lead to deceleration traffic condition. Region (7) represents the more deceleration mode of PCC. In this mode, subject vehicle decelerates more than the calculated rate in ACC mode by reducing the Std_{do}.

In the PCC system, there is a possibility that collision risk can be underestimated compared to ACC system, because the safety of a certain vehicle is only related to its immediate leading vehicle. For example, let us consider the case when the four leading vehicles are proceeding with high speed and only the immediate leading vehicle suddenly stops for some reason. In this case, since the PCC uses a weighted average DSSM value of the multiple leading vehicles, it may suggest less deceleration than ACC. In order to overcome this limitation, we have decided that a single vehicle's Std_{do} should be smaller than 1.0 and that the full deceleration should be made at 1.2 of a single vehicle's DSSM value.
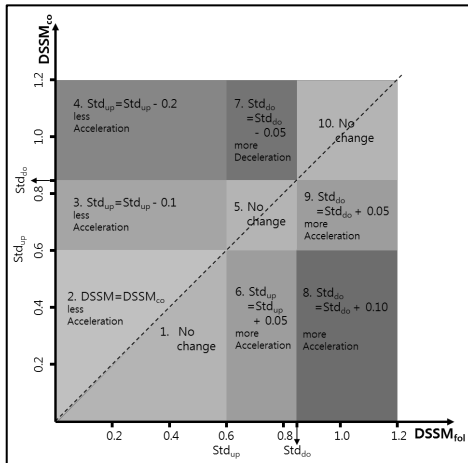
Figure 3. $DSSM_{fol}$ versus $DSSM_{co}$ diagram

Evaluation of the Predictive Cruise Control

To evaluate the proposed control algorithm, we simulate the proposed ACC and PCC algorithms. All cases include one or more sections that experience shockwave. For the simulation, we used the Next Generation Simulation (NGSIM) trajectory data collected from a highway site called I-80 in California, US (2006). The effectiveness of proposed ACC for single vehicle and PCC can be evaluated by verifying whether the control algorithm is able to attenuate the shockwave and reduce the energy

consumption during acceleration. As shown in Figure 4, both ACC equipped vehicle and PCC equipped vehicle show a stable trajectory than human driver. Especially, PCC shows the best capabilities that can attenuate the shockwave because the proposed PCC can predict the shockwave in advance based on the information of the preceding vehicle. As shown in Figure 4, PCC makes a smooth trajectory by reducing the acceleration rate before meeting shockwave. Conversely, PCC also makes a smooth trajectory by reducing the deceleration rate or accelerating before leaving the influential area of shockwave. This smooth trajectory reduces the negative effects of shockwaves.
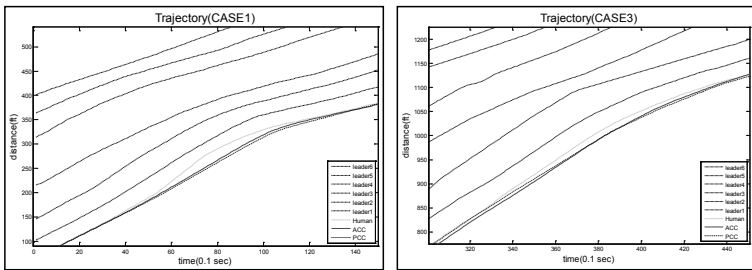
Figure 4. The trajectory of the human driver, ACC equipped vehicle and PCC equipped vehicle.

There are several ways to estimate the energy efficiency. In this paper, the energy efficiency of the each mode is estimated based on the acceleration amount. The acceleration amount is the sum of each time step's acceleration value. As shown in Table 1, PCC generally shows the best energy efficiency, and ACC shows the second higher energy efficiency. Compared to other modes, PCC can effectively remove the unnecessary deceleration and acceleration behaviors by predicting the future traffic condition. For example, PCC reduces the acceleration rate or speed when shockwave is coming even though acceleration is recommended considering only the first leader vehicle. From the perspective of collision risk, DSSM of PCC generally shows the similar trend with DSSM of ACC, because PCC takes deceleration or acceleration action in advance to reduce the collision risk by predicting the future traffic condition.

		Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
Human	Acceleration amount	863	636	746	568	846	813	1099
	Mean DSSM	0.82	0.66	0.78	0.77	0.47	0.51	0.48
ACC	Acceleration amount	519	523	427	552	708	635	742
	Mean DSSM	0.67	0.72	0.73	0.72	0.77	0.78	0.76
PCC	Acceleration amount	470	498	502	461	607	551	604
	Mean DSSM	0.68	0.73	0.75	0.73	0.70	0.77	0.73

Table 1. Acceleration and DSSM of human driver, ACC equipped vehicle and PCC equipped vehicle.

Conclusions and Future work

This paper proposed and analyzed the design of the new Predictive Cruise Control (PCC) systems from the point of view of improving traffic flow stability and

safety performance while reducing the energy consumption of the vehicle. The new measure called Deceleration based Safety Surrogate Measure (DSSM) was developed and applied to the Predictive Cruise Control (PCC). Practical advantages of the PCC were simulated and compared to human driver. The results show that PCC mode can significantly improve the energy efficiency while maintaining the similar collision risk (DSSM).

In order to implement PCC system successfully, other things also have to be considered for further development of the proposed PCC. The proposed PCC model has focused only on the individual vehicles' car-following behaviors. It is suggested to perform further studies considering various situations like lane-changing events. Furthermore, this study is performed under the assumption that the communication environment among vehicles is perfect. It is postulated that there is no error or delay during communication process. It is also suggested that further studies to consider communication related factors, such as error rate, computation time, and transmission speed.

Acknowledgment

Reference

DARBHA, S. & RAJAGOPAL, K. R. (1999). "Intelligent cruise control systems and traffic flow stability". *Transportation Research Part C-Emerging Technologies, 7***,** 329-352.
DAVIS, L. (2004). "Effect of adaptive cruise control systems on traffic flow." *Phys. Rev. E.*
IOANNOU, P. A. & CHIEN, C. C. (1993). "Autonomous Intelligent Cruise Control." *Ieee Transactions on Vehicular Technology, 42***,** 657-672.
IOANNOU, P. A. & STEFANOVIC, M. (2005). "Evaluation of ACC vehicles in mixed traffic: Lane change effects and sensitivity analysis." *Ieee Transactions on Intelligent Transportation Systems, 6***,** 79-89.
NGSIM. (2006). Next Generation SIMulation. URL: http://ngsim.fhwa.dot.gov/.
SWAROOP, D. & RAJAGOPAL, K. R. (2001). "A review of constant time headway policy for automatic vehicle following." *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE.
YEO, H. (2008). "Asymmetric microscopic driving behavior theory." *Ph D in Engineering-Civil and Environmental Engineering, University of California, Berkeley*.
YEO, H., SHLADOVER, S. E., KRISHNAN, H. & SKABARDONIS, A. (2011). "Microscopic Traffic Simulation of Vehicle-to-Vehicle Hazard Alerts on Freeway." *Transportation Research Record: Journal of the Transportation Research Board*.
YEO, H. & SKABARDONIS, A. (2011). "Microscopic fundamental relationships between vehicle speed and spacing in view of asymmetric traffic theory." *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*.
ZHOU, J. & PENG, H. (2005). "Range policy of adaptive cruise control vehicles for improved flow stability and string stability." *Ieee Transactions on Intelligent Transportation Systems, 6***,** 229-237.

Issues in BIM for Facility Management from Industry Practitioners' Perspectives

R. Liu[1] and R. R. A. Issa[2]

[1]Construction Science and Management Program, College of Architecture, 501 W.César E. Chávez Blvd, San Antonio, TX -78207; PH (210) 458-3054; email: rui.liu@utsa.edu
[2]Rinker School of Building Construction, University of Florida, PO Box 115703, Gainesville, FL, USA 32611-5703; PH (352) 273-1152; email: raymond-issa@ufl.edu

ABSTRACT

As Building Information Modeling (BIM) becomes widely adopted by the construction industry, it holds undeveloped possibilities for supporting Facility Management (FM). Some FM information systems on the market claim to address the needs of FM managers. However, are the functionalities provided by the current BIM-based FM software companies those actually required by FM Professionals? What data is required by FM professionals in the operation and maintenance (OM) phases of facilities? The aim of this paper is to clarify the needs of FM professionals for maintenance purposes and to investigate potential areas that can use BIM technology to satisfy such data requirements. A survey was conducted to collect perspectives from industry practitioners for the data requirements for FM in the design and construction phases. The survey results indicated that maintainability considerations should be taken into consideration during the facility design phase. The results address the areas perceived by practitioners that need maintainability consideration in the design phase.

Keywords: Building Information Modeling, Facility Management, Data Requirements

INTRODUCTION

The AEC industry has raised a good deal of interest surrounding the use of BIM for facility management. The opportunities for leveraging BIM for facility operations are compelling, but the utilization of BIM in facility management is lagging behind BIM implementation in the design and construction phases (Akcamete et al. 2010). With the development of BIM, knowledge sharing between facility management and design professionals has become a more realistic possibility. BIM technology is being used in the design and construction phases. However, there is a

need to expand BIM beyond the design and construction stages and to consider using BIM for the life cycle of construction projects.

A study by the US National Institute of Standards and Technology (NIST) showed that the annual costs associated with inadequate interoperability among software systems was $15.8 billion (Gallaher et al. 2004). Two thirds of this cost was incurred as a result of ongoing facility operation and maintenance activities (Shen et al. 2010). The problem is because information at different stages needs to be collected repeatedly. With the technology of BIM, it should be possible to collect information once and use it for the whole project life cycle.

In summary, there is a knowledge and technology gap between design and facility management professionals. The first step for design for management (D4M) is to find out the data requirements of facility maintenance activities. To maximize the benefit of the BIM information database o, the first step is to figure out the data requirement of different parties through the life cycle of the building project. What are the requirements of facility management so that the new technology can bring the right information to the right party at the right time?

LITERATURE REVIEW

The operation phase constitutes approximately 60% of the total lifecycle cost of a facility. The main activities during operations are related to maintenance and repair (M&R). Reactive maintenance and repairs bring excessive expenses. However, the nature of most maintenance work tends to be non-routine and reactive (Mobley et al. 2008). Reactive maintenance and repairs are not efficient, since they cost three to four times more than the planned maintenance for the same repair (Sullivan et al. 2010, Mobley et al. 2008). So it is reasonable to support more planned maintenance work instead of just reacting to failures. Sullivan et al. (2010) recommended among others prioritizing the components in a facility, recording root causes of failures and analyzing equipment failure modes in order to capture reliable information. A reliable maintenance database containing historical information of M&R work is necessary for planned maintenance decisions. Since significant unnecessary expenses occur in current practice, there are ample opportunities for major savings in the operation phase and computerized tools are needed for the improvement of operation and maintenance activities (Akcamete et al. 2010).

Facility managers' involvement in the design phase can be helpful for D4M purpose, but in reality, the facility management team may not yet have been set up during design phase. Thus acquiring the facility management team's general knowledge, such as the data requirements of facility management from the BIM database, which does not require the physical presence of the facility management staff, can be a solution to this problem.

THE SURVEY RESULT

The survey was distributed using a variety of methods: (1) through the Stevens Construction Institute's newsletter, (2) through LinkedIn groups related to FM including: Facility Managers Building Owners Network, Building Owners and

Managers Association International (BOMA), and Integrated Facility Management (IFM), and (3) through an email list, collected by the author, combined with COAA Owners from Higher education, K-12, government and some organizations. It was also distributed to the Florida State University System (SUS) facility management departments. Through March 7, 2012, only 12 complete responses had been received. With the help of the buildingSmart Alliance, the survey was distributed again on March 24, 2012, and through April 13, 2012, there were 693 visits, 22 partial responses and 38 complete responses. Since it is not possible to determine the exact number of people who received the survey link because people who got it had the ability to forward it to anyone they thought appropriate, the response rate can only be calculated based on the number of visits and the number of complete and partial responses. The response rate was (38+22)/693=8.66%, while the effective response rate was 38/693=5.48%. As it is with high possibility that some respondents have received the survey many times because they are in more than one groups that received the survey but they only submit one complete or partial survey, they may visit the survey site many times, as a result, the number of people got the survey should be less than the number of visits. Even we cannot calculate the accurate survey response rate, it should be better than 5.48%. In addition, this survey is still ongoing, this study is still receiving responses.

Part I: Demographic Distribution

Figure 1 shows the distribution of the survey respondents. Architects constituted the largest proportion (8, 21.1%), followed by construction managers and facility managers (5, 13.2%), consultant companies (4, 10.53%), higher education managers (3, 7.9%), owners (3, 7.9%), general contractors (2, 5.3%), design builders (2, 5.3%), two software developers and employees from four other organizations including one laboratory, one specialty subcontractor, one attorney and one research institute.
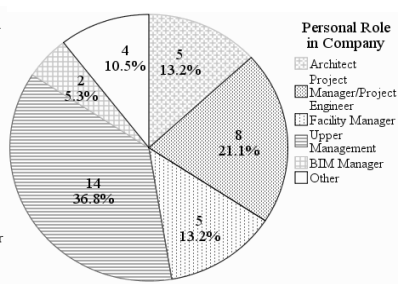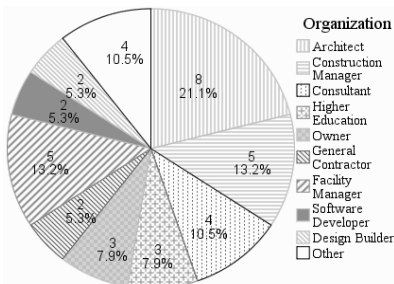
Figure 1. Respondents' organization   Figure 2. Respondents' role in company

Different roles in the same company may have different perspectives and ideas. Figure 2 shows the distribution of the respondents by role in their companies. Upper managers constituted 14 out of the 38 (36.8%) respondents, followed by eight project managers and project engineers (21.1%), five facility managers (13.2%) and five

architects (13.2%). Management level respondents constituted two-thirds of all the responses. Even when the survey was specifically distributed to maintenance related staff and MEP engineers, none of them answered the survey. The reasons behind this may be: first, maintenance staff themselves are not familiar with BIM, so after they read the instructions accompanying the survey, they were not able to respond. Secondly, they may not have had convenient access to the computer or web as management level respondents have.

As shown in Figure 3, the respondents' BIM experience ranged from "no experience" to "expert." Among the 38 respondents, only nine (23.7 %) considered themselves as experts, and five (13.2 %) defined themselves as advanced BIM users. Ten (26.3 %) of them described themselves as intermediate users. Nine (23.7 %) respondents classified their BIM experience level as beginners and five (13.2 %) of the respondents had no experience with BIM projects.
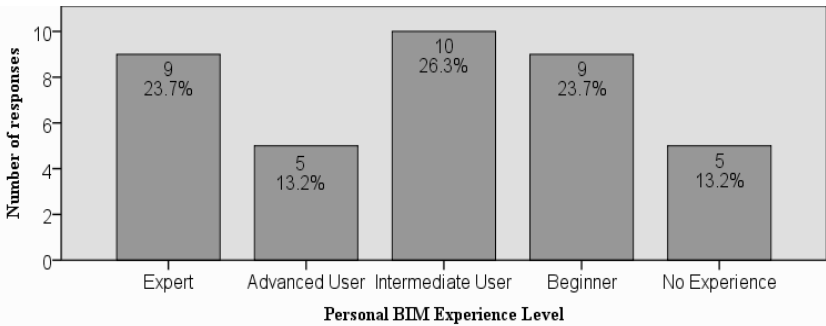
Figure 3. Respondents' personal BIM experience level
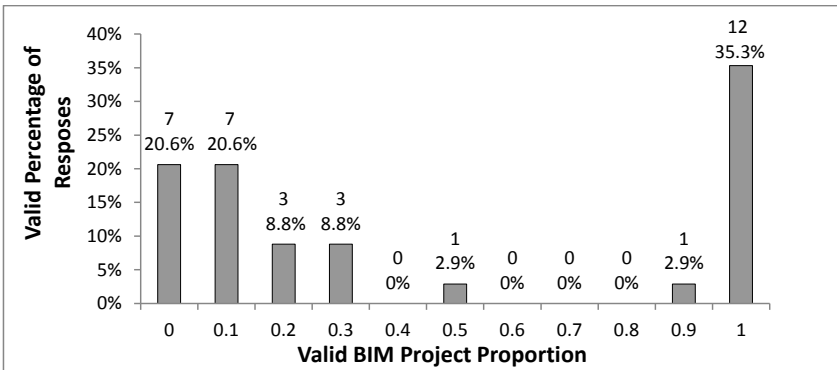
Figure 4. BIM project proportion

In order to determine their objective involvement in BIM projects from the responses, the proportion of BIM projects among new projects in the past 12 months was calculated from the respondents' answers about their number of new projects and the projects that used BIM. As shown in Figure 4, among the 34 effective responses,

seven respondents (20.6 %) had no BIM project for the last year. Seven of them (20.6 %) had less than 10% of their projects that utilized BIM. There were 12 (35.3 %) respondents who utilized BIM in all their new projects. From Figure 4, the BIM utilization of these companies is polarized. The company either uses BIM for all their projects or uses BIM for a very limited proportion of their new projects.

Part II: Perceptions of Data Requirements for BIM-Assisted Facility Management

In the respondents perspectives, the manufactures and/or suppliers did not provide adequate information in regards to maintenance activities in the operation phase. Five out of 31 (16.1%) respondents experienced adequate information, 23 (74.2%) respondents indicated that they received some information from manufactures while four respondents' indicated that they received no information at all from manufacturers.

For the 25 respondents with BIM model in use, two of them responded that the model carried enough information for facility Operation and Maintenance, three indicated that there was no information needed for OM in the BIM model, and all info for OM were manually input. Nine respondents indicated that their BIM models had part of information needed for OM and manual input was still required. Eleven respondents indicated even if they required a BIM model built and required for design and construction, the model was not yet used for the FM phase or that they had difficulty doing so.

Ten out of 24 (43.5%) effective responses showed that the existing work order systems did not provide enough information enough to complete the requested work; fourteen (53.8%) responses indicated that their existing work order system had enough information for requested work. Figure 5 shows the responses to the questions about what specific information is needed in work orders for equipment repair.

Besides the above attributes needed for equipment repair, respondents also listed that connectivity and affected spaces for the out-of-service equipment, geospatial link from the element to the model, date of installation and warranty information, and repair history should be available for equipment repair work.
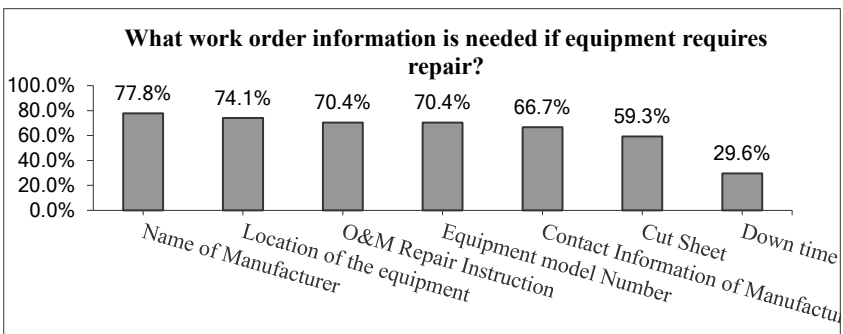
Figure 5. Information needed for work order

In addition, questions were asked about information that is needed for O & M but that is always hard to locate. Based on the respondents' answers the information for facility operation and maintenance that is typically hard to locate includes:

- Equipment operating parameters and spare parts
- MEP information, specifications and warranty
- Electrical panel information
- Make, model, and O&M manuals
- Security and HVAC details
- Work order history information
- Up-to-date as-built plans

As shown in Figure 6, there were only three (9.4%) respondents who indicated no involvement of FM personnel in both design and construction phases, eight (25.0%) responses indicated FM involvement in the design phase, five (15.6%) respondents experience FM involvement in construction phase, and 16 (50.0%) out of 32 responses had FM personnel involved in both the design and construction phases.
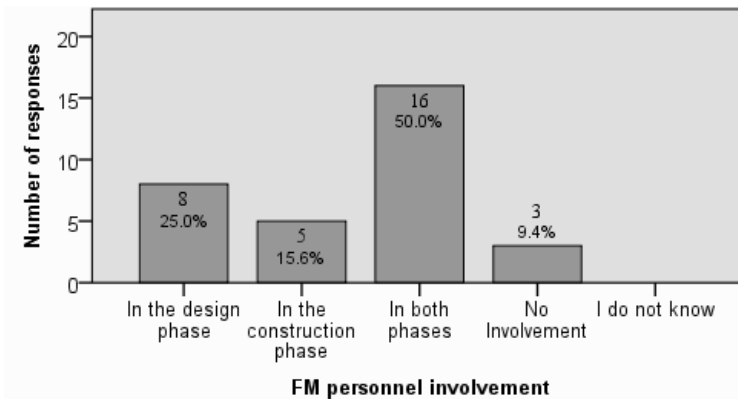
Figure 6. FM personnel involvement in design and construction phases

DISCUSSION OF SURVEY RESULTS

The survey results indicated that industry practitioners believed that maintainability issues should be considered in the design and construction phases and they put forward the efforts to make it happen. As shown in Figure 6, 29 of the 32 respondents had considered maintainability in the design and/or construction phases and half of the respondents indicated FM personnel involvement in both the design and construction phases. However, only 16.1% of respondents expressed that they had adequate information from the manufacturer for the operation phase. Moreover, 10 out of 24 (43.5%) of the respondents expressed that they did not have enough information for the requested work orders and some information that is required for OM but hard to locate from existing documents. These results indicate that even if FM personnel got involved in the earlier phases, the existing business process and

practices cannot provide a channel for the requirement data to be delivered effectively from design and construction phases to OM phase. Better practices and technologies are needed to carry such information with the understanding of FM's needs. As BIM can serve as a database, it is capable to help smooth the information delivery process if the FM's data needs can be clarified.

In addition, the survey had no responses from civil engineers, structural engineers, structural designers and fabricators, MEP subcontractors, MEP manufacturers and MEP suppliers, although there were such companies in the pool. The absence of these organizations can be meaningful. First of all, these groups may not have opportunities to get involved in BIM or facility management and perhaps they are not interested or familiar with the topic studied. Secondly, cooperation in the AECO industry is still segmented and different stakeholders are working in their own areas without sharing technology and information with others. Moreover, as the survey has BIM in its title, people with an interest in BIM were more inclined to answer it than those who did not have knowledge of BIM or no interest in BIM.

Finally, the response rate of this survey is relatively low especially from facility management practitioners even the survey link has been distributed to several facility management groups. The reason for the low response from FM may because the lack of knowledge of Building Information Modeling of our FM people. When FM got the survey, they do not understand the BIM in the title and avoid answering these questions. If this is the reason behind the low response rate, it proves the gap of BIM implementation between the design, construction and FM. In addition, for the expansion of BIM from earlier phases to FM phase, it requires the efforts not only from FM but also all other segments of the AECO industry, the survey is getting attention from all industry practitioners but not only FM group.

CONCLUSIONS AND FUTURE RESEARCH

BIM technology has been used effectively in the design and construction phases. There is a need to expand BIM beyond the design and construction phases and to consider using BIM for facility management such as in maintenance activities. However, the research on BIM use for facility management is lagging behind the study of BIM in design and construction phases. BIM is not merely a 3D model. A real building information model should hold information for different stakeholders at different phases of the facility's life cycle. The need for information that can be inserted, extracted, updated, or modified, is apparent in all phases of the facility O&M. Since BIM can used as a database for information exchange among construction stakeholders, it is necessary to investigate what information is needed for different parties and how they need the information to be presented. Technology development without understanding the requirements of the end user's need will not contribute to improving business processes. This study is designed to investigate the data requirements of FM personnel in order to optimize the BIM benefits for owners and facility managers.

This study is a first effort in bridging the gaps between the design, construction and facility management phases in the AECO industry. The survey results indicate that there are a lot of other issues such as vandalism, commissioning, auditing, and

lack of guidelines for owners, legal problems, which need further study and which are important for realizing the cooperation and sharing of information among different parties of AECO projects. These problems are not covered in this study but are definitely worth addressing in the future research.

REFERENCES

Akcamete, A., Akinci, B., and Garrett, J. H. (2010). "Potential utilization of building information models for planning maintenance activities." *Proceedings of the International Conference on Computing in Civil and Building Engineering* W. Tizani(Editor), June 30-July 2, Nottingham, UK,Nottingham University Press, 151-157.

Gallaher, M. P., O'Connor, A. C., Dettbarn, J. L., and Gilday, L. T. (2004). "Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry." *NIST GCR 04-867* National Institute of Standards and Technology, Gaithersburg, Maryland.

Mobley, R. K., Higgins, L. R., and Wikoff, D. J. (2008). *Maintenance Engineering Handbook*, McGraw-Hill, New York.

Shen, W., Hao, Q., Mak, H., Neelamkavil, J., Xie, H., Dickinson, J., Thomas, R., Pardasani, A., and Xue, H. (2010). "Systems integration and collaboration in architecture, engineering, construction, and facilities management: A review." *Advanced Engineering Informatics*, 24(2), 196-207.

Sullivan, G., Pugh, R., Melendez, A. P., and Hunt, W. D. (2010). "Operations & Maintenance Best Practices: A Guide to Achieving Operational Efficiency." U.S. Department of Energy: Federal Energy Management Program, Wachington, D.C.

Zhang, S., Teizer, J., Lee, J.-K., Eastman, C. M., and Venugopal, M. (2012). "Building information modeling (BIM) and safety: Automatic safety checking of construction models and schedules." *Automation in Construction*. doi:10.1016/j.autcon.2012.05.006

Knowledge Discovery of Spatial Conflict Resolution Philosophies in BIM-enabled MEP Design Coordination using Data Mining Techniques: a Proof-of-Concept

Li Wang[1] and Fernanda Leite[2]

[1]PhD Candidate, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 1 University Station C1752, Austin, TX, 78712; email: liwang@mail.utexas.edu
[2]Assistant Professor, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 1 University Station C1752, Austin, TX, 78712; email: fernanda.leite@austin.utexas.edu

ABSTRACT

Knowledge discovery and formalization in a computer interpretable manner is a critical step toward effective construction automation. Ideally, any spatial conflicts in design should be identified and coordinated prior to construction execution. Building Information Modeling (BIM) has greatly improved the efficiency of spatial conflict identification and coordination, compared to the traditional paper-based approach. However, the current model-based design coordination is still an iterative and experience-driven process, during which considerable knowledge is not properly captured. This research attempts to explore the feasibility of formalizing knowledge regarding spatial conflict analysis in design coordination by capturing necessary information using an object-oriented data capture approach in a model-based environment and analyzing the captured data with data mining techniques. It is envisioned that the formalized knowledge can then be used to narrow down the search space and assist future coordination tasks. Specifically, this paper proposes a process of information acquisition, data capture and knowledge discovery with BIM and supplemented tools. Results of the feasibility study are summarized and discussed. Opportunities and challenges are identified.

INTRODUCTION

Given that the earlier the changes are made the less the cost will be (Paulson 1967), design review and coordination before construction is curial to project success by eliminating constructability issues and ensuring design quality before field installation. Design coordination in Mechanical Electrical and Plumbing (MEP) systems is considered by many construction professionals one of the most challenging tasks in the delivery process of construction projects (Tatum and Korman 2000, Korman et al. 2003). The general concept of MEP coordination involves defining locations and dimensions of MEP components in congested spaces to avoid interference between pairs of disciplines which includes mechanical, electrical, plumbing, structural, architectural and fire protection, while complying with design

and operations criteria (Korman and Tatum 2001, Korman et al. 2003). The advancement of information technology is changing the way people work, think and communicate. Nowadays, Building Information Modeling (BIM) has been widely used in the building construction industry in the United States, mostly for design or trade coordination (Hartmann and Fisher 2007, Becerik-Gerber and Rice 2010). With the assistance of BIM, the construction team can perform automated clash detection to identify clashes between systems more efficiently and intuitively, as compared to paper-based design review (Songer et al. 1998; Staub-french and Fischer 2001, Staub-French and Khanzode 2007, Khanzode et al. 2008, Leite et al. 2011).

However, although the process of identifying clashes has been expedited by formalized knowledge and advanced technology, the process of resolving MEP design conflicts is still very ad hoc, because it requires distributed knowledge from different trades to be integrated and collaborated for decision making (Korman et al. 2003). Although most of the clashes discussed in the coordination meeting had repetitive patterns, the majority of knowledge involved was tacit knowledge based on specialized expertise and experiences, which is difficult, if not impossible, to centralize or formalize. Furthermore, the information used and generated during the design review process was either not documented or not properly documented, and the lessons-learned from the review process was usually implicitly carried away by certain experts rather than shared with the project team for future benefits. The lack of formalized knowledge for MEP design conflict resolution and inadequate historical data available hinders the attempts towards streamlining and expediting the decision making process, and also impedes knowledge reuse and transfer across different disciplines (e.g., between design and construction) and different projects.

We envision that by capturing and analyzing historical data relevant to coordination issues, tacit knowledge of MEP design conflict resolution can be semi-automatically extracted and formalized, which will reduce the reliance on individual researchers and provide efficiency in hidden pattern recognition. This paper proposes a new approach to formalize knowledge and discusses the feasibility, potential benefits as well as the challenges of implementing the proposed knowledge discovery method in the MEP coordination process.

BACKGROUND RESEARCH

Many research efforts have focused on formalizing and characterizing design-relevant constructability knowledge. Fischer (1991) suggested a framework classifying design-relevant constructability knowledge for reinforced concrete steel structures as application heuristics, layout knowledge, dimensioning knowledge, detailing knowledge, and exogenous knowledge. Hanlon and Sanvido (1995) extended this framework to design rules, performance, resource constraints, external impacts and lessons learned. Korman et al. (2003) identified three knowledge domains (i.e. design criteria and intent, construction issues, and operations and maintenance) that are important for MEP coordination and the knowledge items related to each domain, which built on earlier constructability frameworks. The attributes identified by Korman et al. (2003) include geometric characteristics (e.g., coordinate information, component dimensions and number of connections per length) and topological characteristics (e.g., location, spatial relationships and spatial

adjacencies). Based on this framework, Tabesh and Staub-French (2005) further classified the MEP coordination tasks into conceptual reasoning coordination tasks (i.e., design validation, detailing, and sequencing) and spatial reasoning coordination tasks (i.e., layout, routing and positioning) and the underlying reasons behind the constraints identified in each discipline (i.e., tolerance, productivity, space, performance, access, safety and aesthetics). Khanzode et al. (2008) provided a sequence of coordination between different trades as general guidelines. Leite et al. (2009) identified information requirements for MEP clash identification in manual coordination, which were mainly geometric and location data including distance between surfaces, outer diameters of components, location and elevation. Previous research provides an initial list of attributes that may be considered for MEP conflict resolution. However, this list needs to be refined since the focus of these studies was clash identification instead of resolution. Moreover, the approaches used in these studies were top-down approaches (e.g., interviews and case studies) aiming to formalize knowledge by searching for information from answers and descriptions provided by the experts. Top-down approaches are good for in-depth analysis but are difficult to be implemented on a large scale, because it is very time-consuming and expensive to build such expert systems. So far the scope of formalized domain knowledge remains small and static due to limited number of cases investigated. Furthermore, the results generated with top-down approaches can be subjective because the information gathered and result interpretations largely depend on the expertise and skills of the investigator.

As indicated by Conklin and Burgess-Yakemovic (1995), capturing design rationale is a difficult problem if documenting the decisions and why the decisions are made is viewed as a separate process from constructing the artifact. For conflict resolution in MEP design, as long as the conflicts are removed, the project team usually does not care *how* and *why* a solution is generated and implemented; thus, conflict resolution (CR) rationale is rarely formally captured throughout the coordination process. Guidelines of how to capture CR rationale intuitively and how to use the identified rationale to assist future coordination tasks need to be developed.

In the AEC industry, a huge amount of data is generated and circulated in every project and also during MEP design coordination. Based on this fact, data mining can be a promising way for knowledge discovery and continuous improvement. Software developed for design purposes in the 1980s led to early attempts to use heuristics derived from explicit human experience in a limited compilation of constructability knowledge (Kirby and Cannalte 1988; Kirby et al. 1991). Skibniewski et al. (1997) investigated the use of machine learning approaches for constructability analysis. Soibelman and Kim (2002) suggested that the knowledge discovery application might be used to identify time overruns in construction activities by using decision trees and neural networks using resident management data. It has been validated in previous studies that knowledge formalization using historical data has potential strength in the experience-oriented construction industry. The current state of practice for knowledge documentation in MEP design coordination is unstructured and informal. The 3D model does not provide proper documentation template to capture important attributes and decisions in MEP coordination meeting so that the information can be referenced and analysed

for knowledge formalization and knowledge reuse (Staub-French and Khanzode 2007). Currently, large amounts of data related to MEP design coordination are not available and there is no guideline for a structured documentation process, which is a big challenge for using a data-driven approach to formalize MEP design conflict resolution knowledge.

RESEARCH APPROACH

This paper aims at exploring the possibility of identifying the rationale of MEP design conflict resolution by reasoning about relevant historical data. The feasibility study contains a sequence of three steps: attribute selection, data documentation and reasoning.

As the first step, attribute selection includes defining what are the decisions to be made and what information needs to be captured so that it can represent the rationale of decisions made. In MEP design coordination meetings, the most important decision is determining which trade(s) should take the responsibility to make changes so as to avoid conflicts with others. Other decisions include determining what actions should be taken, by which date the changes should be made and what consequential impacts the changes will cause. This feasibility study focuses on the most fundamental decision, which is determining the trade(s) that should respond to resolve the conflict. A list of attributes that describe the coordination issues was identified based on literature review and refined in a case study.

The next step is to determine how to efficiently capture the values of the identified attributes in a model-based environment and how to store the captured data for future reference and analysis. The clash data used in this study was gathered from an ongoing construction project in Southern United States. In this project, 238 clashes were documented, within which 232 clashes were Mechanical, Electrical, Plumbing and Fire Protection (MEPF) related. The geometric data was directly extracted from the building information models. The non-geometric data was captured from drawings, specifications, clash report, meeting notes, as well as observations in coordination meetings and expert interviews. A database was constructed to organize and store the project data.

The final step is to explore different reasoning mechanisms for pattern recognition so as to identify a rationale for decision making. A classification model was developed and the results using different algorithms were compared. Four types of classifiers (i.e., decision tree classifiers, rule-based classifiers, instance-based classifiers and Naive Bayesian classifiers) were examined in the feasibility study; more will be examined in future analysis. The decision tree classifier is a tree-based classification technique that is simple and widely used. Some algorithms include ADTree, BFTree, J48, LMT, RandomTree, REPTree and SimpleCart. The rule-based classifier classifies records by using a collection of 'if...then...' rules. Some algorithms include ConjunctiveRule, JRip, NNge. OneR, PART and ZeroR. The instance-based classifier uses some distance/similarity function to predict the class of an unknown instance based on the class of those training instances similar to it. Some algorithms include IB1, IBk, KStar and LWL. The Naive Bayes classifier uses the Bayes theorem to predict class labels assuming that the input attributes are conditionally independent of each other. Some algorithms include BayesNet,

NaiveBayes, NaiveBayesSimple and NaiveBayesUpdateable. In this study, 22 classification algorithms in these four categories were applied to the datasets using Weka 3.6.8 software which contains open-source machine learning algorithms for data mining tasks.

RESULTS

This section presents the results from the feasibility study following the aforementioned three steps: attribute selection, data documentation and reasoning.

Attribute Selection

Four groups of attributes are selected, which include geometric feature, component feature, functional feature and solution class. In the feasibility study, the solution class is the responding system that is responsible to react and solve the identified problem. Geometric information includes attributes related to the size (e.g., height, width, length, cross-section area), top/bottom elevation, shape and configuration of a model component. Since the focus of this study is spatial conflict resolution, geometric data is most important to consider. Geometric data can be directly or indirectly extracted from the design models. Component feature refers to the non-geometric component-based properties (e.g., the system a component belongs to, material type and cost, or the number of clash points related to a component). These features can be extracted from the model if it is created properly. Functional features may include joint/connection type, supply content, number of connections, clearance, tolerance, gravity/pressure-driven, support system, installation space, installation sequence, installation cost, lead time for fabrication, access space, access frequency, aesthetics and safety. Functional features contain knowledge from design, construction and operation domains and are comparatively difficult to be explicitly included in the model. Table 1 shows a list of simplified attributes used in this feasibility study. A functional feature is presented as 'clash type', which roughly characterizes what kind of clashes it belongs to. Hard clashes refer to the physical conflicts between systems; clearance clashes refer to interference with clearance space; tolerance clashes refer to inadequate space left to account for tolerance; functional clashes refer to the other conflicts that cause problems in construction sequence or hamper systems to function properly or other issues related to maintenance, aesthetic and safety. In future studies, this attribute list will be extended and refined.

Table 1. Attributes Description.

Feature	Attribute	Type	Example
Geometric	Shape	Nominal	Flat, round
	Cross-section area	Numeric	731 (sq in)
	Volume	Numeric	23757 (cu in)
	Available space	Nominal	Y, N
	Clashing volume	Numeric	12 (cu in)
Component	Trade	Nominal	Electrical, plumbing
	Material Flexibility	Nominal	Y, N
	Clash points	Numeric	1, 2

| Functional | Clash type | Nominal | Hard, clearance, tolerance, functional |
| Solution | Respondent | Nominal | A, B, A & B, None |

Data Documentation

Table 2 shows a summary of clashes identified in a construction project categorized by trades. 136 clashes are related to mechanical ducts; 80 clashes are related to plumbing; 76 clashes are related to electrical; and 63 are related to fire protection. The largest number of clashes (46 clashes) was between mechanical ductwork and electrical systems and it is used as an example in the feasibility study.

Attributes described above were documented for each clash using an object-oriented data capture tool which links model elements with review data (Wang et al. 2012). At this stage, data was captured and prepared manually. It is envisioned that the data documentation process would be integrated with current design review procedures and supported by an efficient documentation template.

Table 2.Summary of Clashes.

	MDUCT	MPIPE	PLMB	ELEC	FP	STR	ARC	Total
MDUCT		1	19	46	27	23	20	136
MPIPE			1	1	3			5
PLMB				10	9	37	4	60
ELEC					12	2	5	19
FP						6	6	12
								232

*MDUCT=mechanical duct, MPIPE=mechanical pipe, PLMB=plumbing, ELEC=electrical, FP=fire protection, STR=structure, ARC=architecture

Classification

The class label in this feasibility study is the responding system. Classification in data mining is employing a learning algorithm to identify a model that best fits the relationship between the attribute set and the class label of the input data. The model generated should accurately predict the class labels of input records. Given the difficulty of having experts explicitly and clearly articulate the rules behind the decisions made, the goal of classification is to predict solution classes based on previous decisions made using different learning algorithms. Table 3 shows a summary of the classification results. The algorithms that have the best performance for each data set are listed.

Table 3. Results Summary.

Data Set	Attributes	Classification Algorithms	Precision
MDUCT vs ELEC (46)	Shape, cross-section area, available space, clashing points, material flexibility	*Decision Tree:* J48	81.71%
		Rule-based: DTNB	78.57%
		Bayes: Naïve Bayes	71.43%
MDUCT vs FP (27)	Shape, cross-section area, available space, volume, trade, material flexibility	*Decision Tree:* NBTree	83.92%
		Bayes: Naïve Bayes	81.13%
		Rule-based: DTNB	78.01%
MDUCT	Available space, cross-section	*Decision Tree:* ADTree	85.71%

vs PLMB (19)	area, clashing points, clash type	*KNN:* LWL	85.71%
		Bayes: Naïve Bayes	71.43%
ELEC vs FP (12)	Available space, trade, material flexibility, clashing points, clash type	*Bayes:* BayesNet	80.00%
		Rule-based: NNge	80.00%
		Decision Tree: NBTree	60.00%
ELEC vs PLMB (10)	Cross-section area, available space, volume, clashing points, clash type	*Bayes:* Naïve Bayes	66.67%
		Rule-based: JRip	60.00%
		Decision Tree: J48	60.00%
PLMB vs FP (9)	Cross-section area, available space, trade, material flexibility, clashing points	*Rule-based*: ZeroR	66.67%
		Decision Tree: NBTree	66.67%

*MDUCT=mechanical duct, MPIPE=mechanical pipe, PLMB=plumbing, ELEC=electrical, FP=fire protection, STR=structure, ARC=architecture

Precision was used as a performance metric for result comparison. The evaluation of a classification model's performance is based on the number of the test records it correctly and incorrectly predicts. The datasets with larger amount of data tend to have better results with higher credibility. For example, in the classification between mechanical duct and electrical system, the models that have good performance are the ones generated by J48 (81.7%), DTNB (78.57%) and Naive Bayes (71.43%). Important attributes include shape, cross-section area, available space, clashing points and material flexibility. Although there is an underlining order of preference, the actual situation is quite complicated. The classification models can help deal with the complexity and uncertainty of design coordination. In order to provide more reliable results, additional data will be included in future analysis.

CONCLUSION AND FUTURE WORK

Formalizing coordination knowledge from collective historical data can be a promising approach because it can make use of the past project experience, reduce the subjective impact of individual and provide considerable accuracy of prediction. The information and knowledge derived from project databases can continually informs intelligent decision making and assist future generation of design. With this perspective, this paper indicates that if past MEP coordination data is documented properly, it can be used to make accurate predictions for future issues and conflicts. However, currently, the 3D model alone does not provide this type of documentation. A complementary model-based documentation template needs to be developed to support automated or semi-automated data acquisition. Future work includes database refinement, further examination of reasoning algorithms and validation of results.

REFERENCES

Becerik-Gerber, B. and Rice, S. (2010). "The perceived value of building information modeling in the U.S. building industry", *Journal of Information Technology in Construction (ITcon)*, Vol. 15, 185-201, http://www.itcon.org/2010/15.

Conklin, J. and Burgess-Yakemovic, K. (1995). "A Process-Oriented Approach to Design Rationale." *in Design Rationale Concepts, Techniques, and Use*, T. Moran and J. Carroll, (eds), Lawrence Erlbaum Associates, Mahwah, NJ, pp. 293-428.

Fischer, M. (1991). "Constructability input to preliminary design of reinforced concrete structures." *Technical Report No. 64*, Ctr. for Integrated Fac. Engrg., Stanford University, Stanford, CA.

Hartmann, T. and Fischer, M. (2007) "Supporting the Constructability Review with 3D/4D models." *Building Research and Information*, Vol.35, No.1, 70-80

Hanlon, E., and Sanvido, V. (1995). "Constructability information classification scheme." *Journal of Construction Engineering and Management*, ASCE, New York, NY, 121(4), 337–345.

Khanzode, A., Fischer, M., and Reed, D. (2008). "Benefits and lessons learned of implementing building virtual design and construction (VDC) technologies for coordination of mechanical, electrical, and plumbing (MEP) systems on a large healthcare project." *ITcon* Vol. 13, Special Issue Case studies of BIM use, pg. 324-342, http://www.itcon.org/2008/22.

Korman, T.M., Fischer, M., and Tatum, C.B., (2003). "Knowledge and Reasoning for MEP Coordination." *Journal of Construction Engineering and Management,* ASCE, New York, NY, 129(6), 627-634.

Korman, T.M., and Tatum, C.B., (2001). "Development of a Knowledge-Based System to Improve Mechanical, Electrical, and Plumbing Coordination." *Technical Report No. 129*, Ctr. for Integrated Fac. Engrg., Stanford University, Stanford, CA.

Leite, F., Akinci, B., and Garrett, J. (2009). "Identification of data items needed for automatic clash detection in MEP design coordination." *2009 Construction Research Congress*, ASCE, Seattle, WA, 416-425.

Leite, F., Akcamete, A., Akinci, B., Atasoy, G., Kiziltas, S. (2011) "Analysis of modeling effort and impact of different levels of detail in building information models". In: *Automation in Construction*, 20(5), p. 601–609.

Wang, L. Yi, J.L. and Leite, F. (2012) "Information Capture and Representation of Model-Based Design Review Process in Virtual Environments." *Proceeding of the 12th International Conference on Construction Applications of Virtual Reality (CONVR 2012)*, Taipei, Taiwan, 2012.

Paulson, B. (1976). "Designing to reduce construction costs." *J. Con- str. Div.*, ASCE, 102(4), 587–592.

Songer, A., Diekmann, J. and Al-Rasheed, K. (1998). "Impact of 3D Visualization on Construction Planning." *Proceedings of the 1998 International Computing Congress on Computing in Civil Engineering*, Boston, ASCE, 321-329.

Staub-French, S., and Fischer, M., (2001). "Industrial Case Study of Electronic Design, Cost, and Schedule Integration." *Technical Report No. 122*, Ctr. for Integrated Fac. Engrg., Stanford University, Stanford, CA.

Staub-French, S. and Khanzode, A. (2007). "3D and 4D modeling for design and construction coordination: issues and lessons learned", *ITCON*, 12, 381-407

Tatum, C.B. and Korman, T.M. (2000). "Coordinating building systems: process and knowledge." *Journal of Architectural Engineering*, Vol. 6, No. 4, 116-121.

Tabesh, A.R., and Staub-French, S. (2005). "Case study of Constructability Reasoning in MEP Coordination." In: *Proceedings of the Construction Research Congress*, ASCE, San Diego, April 5-7, 2005.

Grand Challenges in Information Modeling for the Architecture, Engineering, Construction and Facility Management Industry

Fernanda Leite[1], Chris Bogen[2], Jie Gong[3]

[1] Assistant Professor, Construction Engineering and Project Management Program, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E Dean Keeton St. Stop C1752, Austin, TX 78712-1094; PH (512) 471-5957; FAX: (512) 232-3070; email: fernanda.leite@austin.utexas.edu

[2] Computer Scientist, Information Technology Lab, Engineer Research and Development Center, US Army Corps of Engineers; PH (601) 636-3111; email: chris.bogen@usace.army.mil

[3] Assistant Professor, Department of Civil and Environmental Engineering, Rutgers, The State University of New Jersey, 96 Frelinghuysen Road, Piscataway, NJ 08854; PH (848) 445-2881; email: jiegong.cee@rutgers.edu

ABSTRACT

The architecture, engineering, construction and facility management (AECFM) industry has been experiencing many changes since inexpensive networked, mobile computing devices have become ubiquitous. With the rising amount of information and data generated in the life cycle of capital projects, information modeling and data interoperability have become a critical element in design, engineering, construction, and maintenance of capital facilities. Recent advances in Visualization, Information Modeling, and Simulation (VIMS) have the potential to address a number of these pressing challenges. The objective of this vision paper is to identify grand challenges related to Information Modeling. This is a companion paper to two others, which will address the areas of Visualization and Simulation. These papers are primers for the VIMS Grand Challenges Report, to be developed by an expert task force assembled by the American Society of Civil Engineers (ASCE) Technical Council on Computing and Information Technology (TCCIT) VIMS Committee.

INTRODUCTION

The cost of computing has been in steady decline as the cost of virtually every other thing goes up. Natural resources are depleting and human populations are increasing apace. Sustainable development and construction has become an

overarching theme in the Architecture, Engineering, Construction and Facility Management (AECFM) industry. For many, it is not difficult to realize that computing is and will be a key enabler for achieving sustainability without sacrificing our living standard. Indeed, sensors and computers are playing increasingly important roles in today's capital project development. With the wide adoption of mobile computing in the AECFM industry, we have now entered into an era where information and data are ubiquitously generated and distributed, and, consequently, project organizations have been facing information and data that are generated at high velocity, large volume, and in a great variety of formats. With the rising amount of information and data generated in the life cycle of capital projects, information modeling has become a critical element in designing, engineering, constructing, and maintaining capital facilities. Recent advances in Visualization, Information Modeling and Simulation (VIMS) have the potential to address a number of these pressing challenges.

The objective of this vision paper is to identify grand challenges related to Information Modeling. This is a companion paper to two others, which will address the areas of Visualization and Simulation. These papers are primers for the VIMS Grand Challenges Report, to be developed by an expert task force assembled by the American Society of Civil Engineers (ASCE) Technical Council on Computing and Information Technology (TCCIT) VIMS Committee.

CHALLENGES IN INFORMATION MODELING

At the last VIMS Committee meeting, held during the 2012 ASCE International Workshop on Computing in Civil Engineering (IWCCE) in Clearwater Beach, Florida, the 35 VIMS Committee members who were present were asked, in a breakout session, to list their top three VIMS grand challenges.

The Information Modeling (IM) task-group then grouped and sorted the list of identified VIMS grand challenges, culminating in the following five identified IM topics: data format and interoperability, big data sources, formalizing as-built data, modeling for stakeholders across the building lifecycle, and modeling expected building behavior. Each of the aforementioned IM challenges is discussed subsequently.

Data Format and Interoperability

According to an often cited 2004 National Institute of Standards and Technology (NIST) study (Gallaher, et al. 2004), the United States capital facilities industry loses approximately $15.8 billion dollars due to interoperability issues between design, engineering, facilities management, and business process software systems. Since 2004, the BIM standards community has been and continues to actively pursue interoperability issues:

- **2006-2012**: The buildingSMART alliance Industry Foundation Classes (IFC) schema is revised from 2x2 to 2x3 (2006) and from 2x3 to Ifc4 (2x4) Release Candidate 1 (2012) and model view definitions such as the Facility Management Handover (FM Handover) are introduced. Many CAD/BIM software vendors provide exports to the IFC 2x3 design coordination view.
- **2006**: The Organization for Structured Information Standards (OASIS) introduces the Open Building Information Exchange (oBIX) with generic building telemetry standards as well as a standard Web service architecture (OASIS, 2013).
- **2009**: Green Building XML (gbXML) becomes a California public non-profit group and invites software development firms to its advisory board (gbXML 2013).
- **2007-2012**: The Construction Operators Information Exchange (COBie) is introduced as a lightweight, human-readable representation of the IFC FM Handover MVD.   In 2012 COBie is approved as part of the National Building Information Model standard (NBIMS-US). From 2008-2013 annual public "challenge" events are hosted where software vendors demonstrate IFC (design coordination and FM Handover MVD) and COBie exports and the standards conformance of these exports are evaluated (East 2013a).
- **2012-2013**: The Heating Ventilation and Cooling information exchange (HVACie), the Building Automation Modeling information exchange (BAMie), Building Programming information exchange (BPie), Electrical System information exchange (SPARKie), and the Water System information exchange (WSie) are introduced as working (not yet balloted) candidate IFC model view definitions and extensions to COBie[1](East 2013b).

While data format and interoperability has improved since 2004, there are still substantial challenges ahead in the way of adoption and quality control. With regards to adoption, BIM standards have surpassed the scope of information typically delivered in CAD/BIM models. For example, BAMie specifies how to represent building automation system connections, their network addresses, and performance history in IFC. Such information is far beyond the scope of the current tools that deliver IFC files. Such forward-thinking data standards are presented at the risk of becoming obscure before tool vendors and users are ready to use them, and ultimately, all standardization efforts are conducted at the risk of obscurity. For BIM standards, minimizing this risk is especially challenging because of the multi-faceted nature of the capital facility industry (i.e. target adopters) and its stakeholders.

---

[1] These buildingSMART alliance project pages are accessible in the Related Pages section of the COBie Common BIM Models page:

http://buildingsmartalliance.org/index.php/projects/commonbimfiles/

With regards to quality control, the ultimate goal is to empower designers and BIM authors with tools that provide warnings and errors when contractual requirements, model view definition requirements, and business process requirements are violated. While some tools provide such capabilities, the validation rules are either localized to the tool, or provided by the data author. In order for BIM quality control to mature, standard methodologies and rule sets should be developed to ensure tool and author independent results.

The new buildingSMART Alliance mvdXML standard begins to provide a standardized, structured representation for model view definition conformance rules, but detailed content rules are expressed in a format (IFC STEP constraint syntax) not recognized by XML parsing libraries. Thus, there are no ready-to-use validation engines that can apply these rules to an ifcXML file. The technical solution to this problem may be easy to address, but the more substantial adoption challenge remains.

Big Data Sources

Typical CAD/BIM models are relatively large files that are generated from performance-intensive software, and the corresponding CAD/BIM data schema can be vast and necessarily complex. The IFC schema defines over 700 entity types and the combination of architecture, mechanical, electrical, and plumbing models for large facilities (e.g., the Clinic model from the buildingSMART alliance's Common BIM Files project, East 2013b) represent hundreds of megabytes of data. Loading and processing such models in memory for Java applications or XSLT processing demands gigabytes of available memory without even considering the demands of visualization. Further computational resources are required for analyzing, querying, validating, and merging model changes between multiple BIM revisions.

As the scope of facility data is modeled in and integrated with BIM, typical data sources will grow larger and the number of stakeholders interacting with BIM will increase. These changes will also require more BIM-enabled diverse software platforms for viewing/updating data; and this should increase the demand for efficient processing of BIM. A couple of areas where these problems could be addressed are querying and compression.

Experimental efforts such as BIMQL (Mazairac & Beetz 2012) are beginning to provide standard query frameworks for BIM models, but these technologies are in early development and do not elegantly support complex queries. Furthermore, efficient query technology will require indexing and database modeling methods that support geometric, temporal (with respect to project life-cycle), and domain-centric query and report requirements.

Lossless compression algorithms and standard normalization methods can reduce the size of BIM and enable more efficient processing. Solibri's IFC file optimizer provides lossless reduction of IFC file size, but this technology is not based on widely accepted methodologies for normalization – and some may argue that important details

are indeed lost in the normalization. Compressing the 3D contents of IFC models according to application profiles may be substantially more challenging as it requires the normalization of disparate geometry representations, polygon reduction, and validation of results (to ensure an acceptable degree of spatial accuracy). The BIM community faces the challenge of establishing standard BIM normalization methodologies and algorithms that become as familiar as Codd's (1971) Third Normal Form (3NF) is to relational database developers.

Formalizing As-built Data

As-built documentation is an essential set of records, consisting of construction drawings, specifications and equipment location, which are kept for facility management purposes. These documents are constantly being created and modified throughout the life of a project. Two challenges arise from formalizing as-built data. The first challenge related to capturing as-built data, while the second related to formalizing as-built data, based on an existing data model. Both challenges are described subsequently.

Technological advancements have made it possible to generate 3D models to assess as-built conditions for construction monitoring purposes, such as verifying conformance to baseline project schedules and contract specifications. Such advancements include laser or image-based 3D imaging technologies. Laser scanning has become the mainstream as-built condition capture technologies. Static terrestrial laser scanners have been widely used in capital project development. Many of these scanners can capture millions of points in a short amount of time. These points can be used to create digital representations of as-built condition of facilities. These digital representations could be in the form of surface models, solid models, or building information models. Mobile terrestrial laser scanners provide mobility to static scanners, which could be essential in capturing as-built conditions for large facilities or infrastructures. On the other hand, abundant computing powers have revived interests in 3D reconstruction from un-calibrated images. 3D reconstruction from a large number of un-calibrated photos using the structure from motion (SFM) approach requires expensive and large-scale matrix manipulation and computation, which used to be only possible with supercomputers but now is in the grasp of personal computers. Due to the cost difference between a digital camera and a terrestrial laser scanner, the possibility of generating dense and accurate point clouds from photos is an exciting development in the AECFM industry (e.g., Brilakis et al. 2010; Golparvar-Fard et al. 2010; Bhatla et al. 2012). As the 3D data capture technology quickly evolves, the battlefront of formalizing as-built data is on generating as-built models in a quick and cost-effective manner (Tang et al. 2010). The data captured by laser scanning or photogrammetric methods typically include point clouds and photo data. Derivation of as-built models such as building information models out of these raw data is often a complicated and labor intensive process. The cost of data modeling is often more

expensive than the cost of field 3D data collection. It is rather ironic that we are getting increasingly better at discretizing the world using technologies but are significantly lagging behind in assembling a discretized world into a holistic one. The latter presents a great challenge to the AECFM community.

Modeling for Stakeholders across the Building Lifecycle

Besides the challenge related to model interoperability to enable data sharing during the project lifecycle, it is important to understand who will be using the model at various stages and what information each stakeholder will need/use, as well as modeling how that facility is expected to perform. An associated challenge is the definition of levels of detail in a model for different operational purposes (Leite et al. 2011). Two current trends in BIM include using the model in the field (i.e., during construction) as well as for facilities management. Applications for BIM to the field include quality assurance/quality control (QA/QC). In both applications, it is necessary to understand who will be using the model and for what specific purposes.

When the facility enters its operation phase, the BIM which was previously used for construction, and updated to its as-built status, can then be transferred to facilities management. Besides including the physical characteristics of the facility, a BIM for facilities management needs to be augmented with service, maintenance and cost information. The model should also include information about objects within the building, such as lifts, ventilation and fire systems, and the relationship between them, in a single repository.

To realize the full value of BIM, data sources must become useful to applications used by a variety of facility managers and occupants (i.e. become more than resources for architecture and structural design software). To illustrate the desired outcome, consider a brief history of geographic information systems (GIS). In the beginning, GIS was a tool for scientists and engineers, and now it has become ubiquitous technology in Google Maps, smartphone navigation, and social networking mash-ups. Realizing this level of success for BIM is a formidable challenge because it may require successful execution of all other challenges presented in this paper.

Modeling Expected Building Behavior

There are many buildings which have not met their designed performance criteria. In many of these cases, owners start to question the validity of building design. In reality, many of these failures have been caused by the ill consideration of occupant behavior. As of now, great efforts are often spent on modeling the building systems and conducting simulations to predict building performances. Modeling building occupant behavior is an equally important aspect in building design. However, much less has been done on modeling building occupant behavior as compared to modeling building systems. This is further complicated by the fact that there is virtually no standard dataset for building user behaviors. Uniform building user behavior is often assumed in the building design process. Accurately modeling building occupant behavior is a

challenge that demands attention from the AECFM industry. To address this challenge, better and finer scale user behavior data need to be harvested and modeled in different geographic zones.

It is as important to consider the expected building behavior in the context of its intended use. During the early planning stages and space programming, the functional requirements of a facility are defined. During occupancy various data monitoring devices collect energy consumption data. In practice, the monitored energy consumption is not compared to the expected use in terms of the intended use of its spaces/zones. The research and development (R&D) community is only beginning to answer these challenges, and conducting meaningful case studies will require a substantial coordination of cross-domain resources.

CLOSING REMARKS

This paper identified five grand challenges in Information Modeling for the AECFM industry, which has been experiencing many changes due to advances in information technology and the increasing amount of digital information that is being collected and used throughout the facility lifecycle. These challenges are: 1) *data format and interoperability* to enable data sharing during the project lifecycle, especially with regard to adoption and quality control of data standards; 2) *big data sources* shared across the facility lifecycle, will require more BIM-enabled diverse software platforms, increasing the demand for efficient BIM processing; 3) capturing and *formalizing as-built data*, including assembly of discretized information into a holistic set representing our physical built environment; 4) *modeling for stakeholders across the building lifecycle,* understand who will be using the model at various stages and what information each stakeholder will need/use, as well as modeling how that facility is expected to perform, in order to take enable ubiquitous use of BIM, like what has been done with GIS/Google Maps; and 5) *modeling expected building behavior*, especially occupant/user behavior. This paper is a primer for the VIMS Grand Challenges Report IM section, to be developed by an expert task force assembled by the ASCE VIMS Committee.

REFERENCES

Bhatla, A., Choe, S. Y., Fierro, O., and Leite, F. (2012). "Evaluation of accuracy of as-built 3D modeling from photos taken by handheld digital cameras." *Automation in Construction*, 28, 116-127.

Brilakis, I., Lourakis, M., Sacks, R., Savarese, S., Christodoulou, S., Teizer, J., and Makhmalbaf, A. (2010). "Toward automated generation of parametric BIMs based on hybrid video and laser scanning data." *Advanced Engineering Informatics*, 24(4), 456-465.

buildingSMART alliance (2013). Model View Definition Page. http://www.buildingsmart-tech.org/specifications/mvd-overview (accessed 1/16/2013)

Codd, E.F. "Further Normalization of the Data Base Relational Model." (Presented at Courant Computer Science Symposia Series 6, "Data Base Systems," New York City, May 24th-25th, 1971.) IBM Research Report RJ909 (August 31st, 1971). Republished in Randall J. Rustin (ed.), Data Base Systems: Courant Computer Science Symposia Series 6. Prentice-Hall, 1972.

East, E. William. (2013a). Construction Operators Information Exchange (COBie) bSa Project Page. http://www.buildingsmartalliance.org/index.php/projects/activeprojects/25 (accessed 1/16/2013)

East, E. William. (2013b). COBie Common BIM Models bSa Project Page. http://buildingsmartalliance.org/index.php/projects/commonbimfiles/ (accessed 1/16/2013)

Gallaher, M.P., O'Connor, A.C., Dettbarn Jr., J.L., and Gilday, L.T. (2004). "Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry." Gaithersburg, Maryland, National Institute of Standards and Technology: 210.

Golparvar-Fard, M., Bohn, J., Teizer, J., Savarese, S., and Pena-Mora, F. (2010). "Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques." *Automation in Construction*, 20(8), 1143-1155.

gbXML (Green Building XML) (2013). About gbXML. http://www.gbxml.org/aboutgbxml.php (accessed 1/16/2013)

Leite, F., Akcamete, A., Akinci, B., Atasoy, G., Kiziltas, S. (2011) "Analysis of modeling effort and impact of different levels of detail in building information models". In: Automation in Construction, 20(5), p. 601–609.

Mazairac W., Beetz, J. "Towards a Framework for a Domain Specific Open QueryLanguage for Building Information Models." (2012). International Workshop on Intelligent Computing in Engineering. July 4-6, Munich, Germany.

OASIS (Organization for the Advancement of Structured Information Standards). (2013) oBIX Technical Committee Homepage. https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=obix (accessed 1/16/2013)

Tang, P., Huber, D., Akinci, B., Lipman, R., and Lytle, A. (2010). "Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques." *Automation in Construction*, 19(7), 829-843.

Cognitive Design of Safety Supervision Instruction

Fernando A. Mondragon Solis, S.M.ASCE[1], William J. O'Brien, Ph.D., M.ASCE[1]

[1]The University of Texas at Austin, Department of Civil, Architectural and Environmental Engineering, 1 University Station C1752, Austin, TX 78712-0273; PH (512) 471-4638; FAX (512) 471-3191; email: {fernando.mondragon, wjob}@mail.utexas.edu

ABSTRACT

Traditional education programs often present work aspects such as safety, quality and schedule as separate entities. Cognitive models of expert field managers are capable of integrating these aspects in terms of the information used in work activities. This paper presents a design of safety supervision instruction based on these models. Such instructional design makes use of specific information contained in the models that is not available through traditional instruction. To develop this approach, the knowledge content of the models is compared to standard safety documentation, in order to identify the different facts and strategies that they posit for safety supervision practice. Then, these differential elements are used throughout the instructional design process to create programs and information tools for construction management education. In this way it is possible to create an instructional environment in which to observe the impact of experts' knowledge on learners' cognitive strategies and use of information.

INTRODUCTION

Construction education is commonly structured as a set of individual concepts and factors whose mutual relationships are not made explicit. This results in disaggregated classroom activities and topics that are not visibly linked to the actual context of jobsite work. The lack of an adequate context may negatively affect the effectiveness of education and training, as expertise is tightly related to the specifics of the application setting (Clark, 2008). These conditions suggest the need to present knowledge in a way that more closely supports the skills that are necessary in construction sites.

This paper presents the use of cognitive work models for construction safety education. The model developed by Mondragon and O'Brien (2012a) thoroughly documented the goals, decisions and information of superintendents at the jobsite. This model points at the information that is necessary for work in specific instances. Furthermore, Mondragon and O'Brien (2012b) also studied the design of learning modules for field management education based on these cognitive models. In such study, elements of cognitive work were utilized to structure and develop an instructional program. This paper expands on such approach to point at the information and strategies of experts practitioners, which are not contained in traditional instruction programs, but can be delivered through cognitive-based instruction.

Two main aspects are covered in this paper: First, the expansion of existing safety curriculum from to the knowledge of expert practitioners; second, the design of

information artefacts from the results of cognitive analysis. To cover these aspects, the following sections of the paper include a background review of current safety curriculum and existing efforts of cognitive-based instruction. Following sections describe the research methodology and current results. The last section presents the conclusion and a discussion of next steps and future research.

LITERATURE REVIEW

Standard Training Program

Compliance with OSHA regulations is a major concern in construction safety (Hinze, 1997). These regulations point at specific rules and knowledge about construction elements and processes that must be followed. Examples include regulations about scaffolds, use of power and hand tools, and communication of hazards on site. In turn, training programs that are based on OSHA regulations are concerned with the use of rules and knowledge, in order to maintain safety practices on site. Such is the case of OSHA-30 training. This type of training is geared towards workers with some safety responsibility on site, as opposed to craft workers strictly. The 30-hour course is usually presented by a facilitator and supported with digital presentations. The emphasis of the program is on hazard identification, avoidance, control and prevention (OSHA, 2011b).

Two limitations can be recognized for this type of training program. One is that the traditional delivery format limits the amount of detail included in the training, which makes it difficult to provide a context for safety activities. The second limitation is that this training does not account for some aspects of work that are known to impact safety performance, such as the schedule of a task, or specific approaches to plan and manage jobsite safety.

Use of Cognitive Models for Instruction

Some studies in construction have looked at ways to include tools and physical models into instructional safety programs. Examples of this are tridimensional game environments (Lin et al., 2011) and simulations (Rojas and Mukherjee, 2006), which can provide richer content and detail than traditional text-based formats. Such strategies present learners with a context in which they can solve safety problems.

Furthermore, other studies have looked into the development of safety models to describe which factors influence safety performance. Some researchers have adopted cognitive approaches that can more adequately consider the complexity of construction organizations (Saurin et al. 2008). For this reason, there has been some interest in the development of cognitive models, in order to account for factors that affect safety performance, such as work design, resources allocation and team processes beyond accident prevention (Mitropoulos, 2009). However, the implications from these models have not been applied to develop instructional programs.

Similar issues have been presented in other domains, where the conventional training programs do not contain the information that is necessary to complete certain work processes. One approach to document the information that is necessary for practice is the use of cognitive analysis techniques. These techniques are useful to

document work from a practitioner's perspective in terms of decisions and information. The representations obtained through these techniques can then be used to develop training and instructional aids. For example, Sullivan et al. (2007) used cognitive task analysis to develop a curriculum for teaching a specific type of surgery to residents. In this study, a cognitive task analysis was developed for a set of experts, to document their main decision points during the surgery and relevant information throughout. Such documentation served to develop training material for surgery students in a way that responded to experts' approach to the surgery—which is different than the strategy presented through the traditional curriculum. This material improved students' cognitive strategies and use of technical knowledge, as compared to students learning from the content of the traditional curriculum. As such, this study shows a way to integrate experts' knowledge from cognitive models into an alternative instructional program.

METHODOLOGY

This research aims to identify and use safety information that is necessary for work and unavailable through standard training. To such purpose, the research is developed in three parts. First, traditional safety curriculum is compared to the content of cognitive analysis documentation, to identify differences in the approach to practice that each source of knowledge indicates. The second part of the research is centered on the development of instructional programs and tools to deliver safety supervision knowledge. Such program is designed to accommodate the differential elements of knowledge found in the first part. In the third part of the research, students participate in the instructional program. This part enables an observation of how learners are able to utilize the information from the cognitive-based material for making decisions and approaching safety supervision practice.

USE OF COGNITIVE-BASED KNOWLEDGE AND STRATEGIES

Comparison to Traditional Training

For the construction industry, OSHA-30 training is divided in two parts (OSHA, 2011b). The first part includes, among other topics, the management of safety and health at the jobsite and the main hazards in the construction industry. The topic of management of safety and health covers jobsite inspections, worksite analysis, and conducting safety meetings. The topic of main hazards in construction presents specific rules for identifying and preventing falls, electrocution, struck-by (e.g., falling objects) and caught-in (e.g. trench hazards) (OSHA, 2011a). Then, the second part of the training presents general topics related to construction, such as scaffolds or excavations. In this part, additional rules are presented—such as the dimension of scaffolds or the height of trench ladders—and also discussions on what are adequate conditions to avoid hazards in regards to each topic (OSHA, 2011a). In this sense, OSHA-30 is structured as a thorough list of unsafe physical conditions that may be found in construction sites.

The cognitive analysis documentation of field managers describes the safety supervision process in a different manner. Among all the decisions registered for field managers, Mondragon and O'Brien (2011, 2012b) identified several decisions related to safety supervision. Figure 1 shows the knowledge representations for one of such

decisions. The lower part of the figure presents the decision (labeled CWR-P12.5), which is to monitor workers' safety practices with respect to safety requirements. This decision happens as part of the goal of supervising field work; to make this decision, a practitioner needs several pieces of information (labeled IRR-P12.5.n), including data of safety trends, OSHA specifications for the tasks being inspected, the work procedures for the task, indications from the job hazard analysis, and an assessment of the capacity of the subcontractors working on a task to complete their schedule. According to the ACWA documentation, that last piece of information is the result of a different decision—deciding if subcontractors' have the capacity to complete schedule—which requires additional information related to the schedule and crew size. This decision is made in the context of a different activity, which is assessing subcontractors' progress—this points at which decisions practitioners need to make as part of planning for safety supervision.
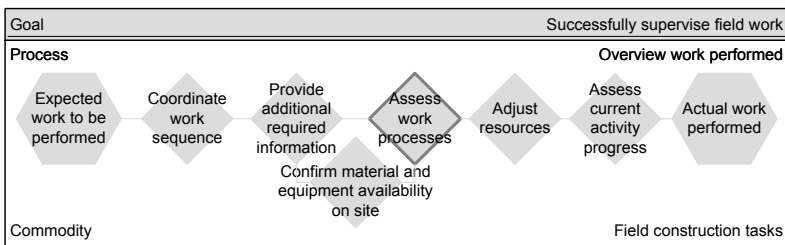
CWR-P12.5 Monitor workers' safety practices with respect to safety requirements for each activity and the project
IRR-P12.5.1 Safety trends and recurrent issues for current work conditions
IRR-P12.5.2 OSHA specifications for the task
IRR-P12.5.3 Work procedures for the task
IRR-P12.5.4 Job Hazard Analysis
IRR-P12.5.5 Subcontractors' capacity to complete schedule

Figure 1. Example of a safety supervision-related decision and information according to cognitive analysis

From the previous analysis it is possible to observe (a) which information is used during safety supervision, and (b) what is the strategy followed by practitioners to plan and execute safety. While OSHA-30 considers some of these aspects, for the most part it is focused on the rules for determining potential hazards given the physical layout of the jobsite. In terms of instruction, this implies that presenting such additional knowledge would enable more realistic evaluations of safety practices, but it would also require an instructional design that could actually guide learners through the experts' work strategies. The next section explains the approach taken in this research to integrate cognitive knowledge into a safety instructional program.

Instructional Design

The approach to bring the elements of cognitive analysis into an instructional program is twofold. First, it is necessary to consider an instructional design that can support the inclusion of components that are meant to provide rich detail about jobsite conditions—such as schedules and activity details. Second, it is necessary to

develop instructional tools that can provide learners only with the information necessary to make decisions—not more information, as it might overload their processing capacity, nor less information, as it might overburden their memory. To such purpose, the EDDE framework was used (Nguyen et al. 2012). This framework compiles instructional principles and interface features to support decisions for technology-assisted instruction in construction management. Recommendations from the framework pointed at skill-building simulations as a learning strategy; this responds to the instructional goal of learning strategies and technical knowledge. The simulation is supported by an array of technologies, including tablet PCs to present multimedia content, and sensors to emulate jobsite conditions and indicate when content should be presented. The EDDE framework also pointed at specific interaction features for supporting technologies, such as simplicity of the interface and inclusion of suggestions and reminders. The simulation has been described in Mondragon and O'Brien (2012b), as well as the development of instructional tools. This paper presents the development of effective content for those instructional tools. Such content is obtained through advancement in the development of the ACWA technique, which points at the inclusion of specific information to support practitioners' decisions in particular contexts.

The ACWA technique, as presented in Elm et al. (2003), can be divided in two parts, one that covers the cognitive analysis of practitioners, and a second one that is concerned with the design of effective user interfaces. Current ACWA studies for superintendents only present the first part (Mondragon and O'Brien, 2012a). For the second part, additional components must be developed to create a design based on the information and decisions obtained. These components, called Representation Design Requirements (RDRs), define the goal and scope of information presentation in terms of the cognitive tasks they are intended to support (Elm et al., 2003). For instance, given a specific work instance (e.g. monitor safety performance during field supervision), RDRs indicate which information should be presented for practitioners and how it should be presented. As an example, RDRs are developed for the decisions and information presented in Figure 1; such development is presented in Figure 2. The RDRs for the decision of monitoring safety practices (Figure 2a) indicate the need to present specific information, such as the trends for safety violation incidents and a means to evaluate OSHA specifications. In turn, these design requirements can be turned into artefacts, such as lists, documents, sketches, tables, forms and schedules. Figures 2b and 2c present two examples of RDR-based designs. Furthermore, one of the information requirements points at another decision, monitoring subcontractors' capability to complete their scheduled tasks. Then, one of the RDRs suggests the use of information concerning the schedule and other details for each task; this is presented in Figures 2d-f.

The artefacts that are produced according to the RDRs were incorporated as part of the instructional program. For example, the table of safety trends in Figure 2b is presented to students prior to performing the field supervision exercise, so they can do a brief analysis of hazards and use this information for supervision; schedule details can be provided during the supervision to aid students decide whether or not subcontractors are running late in a way that can affect safety. The artefacts for the supervision exercise were included in the tablet PC.

Figure 2. Sample of artefacts corresponding to relevant Representation Design Requirements (RDRs)

Participation in the Simulated Environment

Future work for this research considers the participation of civil and architectural engineering students to use the safety module. The simulation will enable individual students to make decisions in realistic jobsite conditions. In particular, it is of interest to measure the extent to which learners can utilize technical knowledge and cognitive strategies. To such purpose, the learning module is divided in three sections. In the first section, a participant is presented with a scenario; in this

part, a specific problem is laid out (e.g. determining if the crew size for a given task can be increased), and technical knowledge is presented so that students can plan for the exercise. The knowledge presented corresponds to the indications of the cognitive model, and it includes information such as OSHA regulations and details about expected jobsite conditions. In the second section, the participant enters the jobsite simulation to evaluate hazards for actual jobsite conditions. Here it is possible to record and measure which information students are able to utilize from the cognitive model to make decisions about onsite safety. In the third and last section, the participant is asked about the effect of their supervision decisions on the overall safety and production performance of the project. Students are also asked to make a decision about the specific problem presented in the scenario, and the information that pointed them to such decision. At the end of this section it is possible to measure the steps of experts' cognitive strategies that students are able to follow.

DISCUSSION

The approach presented in this paper relies on advanced learning strategies and technologies. Although initial setup may require a considerable use of resources, once the technology infrastructure is in place it can be used to support other aspects of construction beyond safety. A materials management learning module has been developed previously (Nguyen et al., 2012), and the safety module can be deployed with the same equipment. Modules can be developed for other activities using the same approach presented here, such as craft supervision or quality control. Future research can focus on: (1) developing other modules; (2) having industry participants; (3a) increasing the number of simultaneous participants to enable teamwork and collaboration decisions; (3b) integrating modules to have learners play different roles at the same time; and (4) experiment with measurements of decision performance, to observe specific improvements in effectiveness of decisions completion, as well as reductions in time between novice and expert positions.

CONCLUSIONS

The sequence of analysis used in this research identifies the differences between existing curricula and practitioners' knowledge, and then structures this knowledge in a way that learners can access it and use it for making decisions. Through the specific pieces of information gathered from experts' knowledge it is possible to integrate different construction aspects into a single activity. In this way this research points at knowledge required for practice that is unavailable through standard training documents. Also, by looking into alternative instructional strategies and media, this paper shows how it is possible to incorporate this information into the classroom. The last part of the research, which concerns participation in the simulation environment, is still undergoing. This part is expected to show how students can use the information presented, as well as provide lessons learned for the instructional strategy and tools.

REFERENCES

Clark, R. C. (2008). *Building Expertise:Cognitive Methods for Training and Performance Improvement* (3rd ed.). San Francisco, CA: Pfeiffer.

Elm, W., Potter, S., Gualteri, J., Roth, E., & Easter, J. (2003). Applied cognitive work analysis: a pragmatic methodology for designing revolutionary cognitive affordances. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 357-382). Mahwah, NJ: Lawrence Erlbaum Associates.

Hinze, J. W. (1997). *Construction Safety.* Upper Saddle river, NJ: Prentice-Hall.

Lin, K.-Y., Son, J., & Rojas, E. M. (2011). A Pilot Study of a 3D Game Environment for Construction Safety Education. *ITcon, 16*, 69-84.

Mitropoulos, P., Cupido, G., & Namboodiri, M. (2009). Cognitive Approach to Construction Safety: Task Demand-Capability Model. *Journal of Construction Engineering and Management, 135*(9), 881-889.

Mondragon, F., & O'Brien, W. J. (2011). Using Applied Cognitive Work Analysis for a Superintendent to Examine Technology-Supported Learning Objectives in Field Supervision Education. *Proceedings of the 2011 ASCE International Workshop on Computing in Civil Engineering, June 19-22*, (pp. 858-866). Miami, FL.

Mondragon, F., & O'Brien, W. J. (2012a). Cognitive Analysis of Field Managers. *Proceedings of the 2012 Construction Research Congress, May 21-23*, (pp. 643-649). West Lafayette, IN.

Mondragon, F., & O'Brien, W. J. (2012b). Cognitive Design of Learning Modules for Field Management Education. *Proceedings of the 2012 ASCE International Conference on Computing in Civil Engineering, June 17-20*, (pp. 97-104). Clearwater Beach, FL.

Nguyen, T., Mondragon Solis, F. A., & O'Brien, W. J. (2012). EDDE: A framework to Explore, Design, Develop and Evaluate Technology-Assisted Instruction for Construction Engineering and Management. *ITcon, 17*, 434-464. Retrieved from http://www.itcon.org/2012/29

OSHA: Ocupational Safety and Health Administration. (2011a). *OSHA 10-Hour Construction Industry Outreach-Trainer Presentations.* Retrieved January 10, 2013, from Occupational Safety and Health Administration: http://www.osha.gov/dte/outreach/construction_generalindustry/const_outreach_tp.html

OSHA: Ocupational Safety and Health Administration. (2011b). *Outreach Training Program Construction Industry Procedures*. Retrieved January 10, 2013, from Occupational Safety and Health Administration: http://www.osha.gov/dte/outreach/construction/construction_procedures.html

Rojas, E. M., & Mukherjee, A. (2006). Multi-Agent Framework for General-Purpose Situational Simulations in the Construction Management Domain. *Journal of Computing in Civil Engineering, 20*(3), 165-176.

Saurin, T. A., Formoso, C. T., & Cambraia, F. B. (2008). An Analysis of Construction Safety Best Practices From a Cognitive Systems Engineering Perspective. *Safety Science, 46*(8), 1169-1183.

Sullivan, M. E., Brown, C. V., Peyre, S. E., Salim, A., Martin, M., Towfigh, S., & Grunwald, T. (2007). The use of cognitive task analysis to improve the learning of percutaneoustracheostomy placement. *The American Journal of Surgery, 193*, 96-99.

Use of Building Information Modeling in Aging-in-Place Projects: A Proof of Concept

Wei Wu[1] and Emily Handziuk[2]

[1]Department of Construction Management, Georgia Southern University, P.O. Box 8047, Statesboro, GA 30460-8047; PH (912) 478-0542; FAX (912) 478-0542; email: wwu@georgiasouthern.edu
[2]Interior Design, Georgia Southern University, P.O.Box 8034, Statesboro, GA 30460-8034, PH (404) 862-0558; FAX (912) 478-5594; email: eh00810@georgiasouthern.edu

ABSTRACT

This research looked at potential use of building information modeling (BIM) in *aging-in-place* (AIP) projects. It hypothesized that BIM could lend a leverage to professionals from the design-for-aging (DFA) community to address the unique challenges of the explosive senior population in the United States. According to the U.S. Census Bureau statistics, more than 80 million citizens will be over age 65 by 2050, a majority of whom prefer to age in place as suggested by a recent national survey. The centerpiece issue in promoting aging in place is to create the physical and service environment catering to the needs in their senior life. To investigate the feasibility and potential benefits of using BIM in AIP projects, a proof of concept was developed through a 360-degree assessment. The major measures used by the assessment included: 1) a series of stakeholder perception surveys; and 2) a prototype model created with common BIM applications. The proof of concept identified typical challenges and concerns of present AIP practices, such as client buy-in, design conceptualization and communication, performance configuration and code compliance. A prototype model was created with the intention to demonstrate practical solutions that BIM could provide to mitigate these challenges and concerns in different use cases and scenarios. Results of the research attested that appropriate BIM integration could streamline the delivery of AIP projects and increase client satisfaction, which would eventually invigorate the AIP effort at large.

INTRODUCTION

Building information modeling (BIM) is gaining rapid uptake in the Architecture, Engineering, Construction, Owner and Operator (AECOO) industry, and seems to continue the momentum (McGraw-Hill Construction 2008). BIM users start to recognize and capitalize from its implementation at different levels across various project types (McGraw-Hill Construction 2009). Nevertheless, under the media spotlight are often those large, striking commercial or healthcare projects, while little attention has been paid to BIM adoption in the residential sector. It may be accounted for by that residential projects appear to be less technical challenging

neither does a single residential project make significant economic impacts, especially when the whole housing market remains struggling. Nevertheless, it should not be denied that the residential sector might as well be able to reap tangible benefits through the leverage of BIM. This research intends to develop such a proof of concept with exclusive focus on *aging-in-place* (AIP) projects. This paper presents the initial results of the project titled as: "Design-for-Aging (DFA) – Investigate the role of BIM in aging-in-place (AIP) projects".

BACKGROUND

Aging in place is the ability to live in one's own home and community safely and independently as one ages (Morley 2012). It utilizes a diverse range of programs that seek to retain senior citizens as integral and productive members of their communities (Ball 2004). In the U.S., aging in place is a national initiative to address the challenges of a rapidly increasing senior population, of which a major constituent is commonly known as the "baby boomers". The impetus to strategically plan for aging in place is statistically justified (Olshansky et al. 2009). According to the U.S. Census Bureau (2012), 83.7 million (21%) U.S. citizens will be over age 65 by 2050. A recent research conducted by the National Conference of State Legislatures (NCSL) and the American Association of Retired Persons (AARP) Public Policy Institute found out that nearly 90 percent of the senior population wanted to age in place, and 80 percent believed their current residence was where they would always live (Farber et al. 2011). Aging in place is a complex geographical process mediated by institutions and other social forces (Cutchin 2003). There are various models used at federal, municipal or community level to promote aging in place, such as the commonly known Adult Day Centers (ADCs) and Assisted Living Residences (ALRs), the very successful Program for All-Inclusive Care (PACE), and more recent development of the Naturally Occurring Retirement Community (NORC) model as well as the Village model (Cutchin 2003, Greenfield et al. 2012, Morley 2012).

At the heart of aging in place, three critical issue areas need to be addressed:
- Healthcare: integration of healthcare delivery with housing and planning initiatives;
- Environment: housing and urban design;
- Planning and zoning: housing stock and location (Ball 2004).

Apparently, "housing" is a centerpiece of aging in place, as human existence is closely related to architectural space and the interactions (visualization and usage) with architectural space supply meaning in the aging process (Cutchin 2003). Architecture can create supportive environments for older people (Devlin and Arneil 2003) with therapeutic effect (Culter 2007). Rioux and Werner (2011) also indicated that the location, accessibility, size and usability of the physical home constituted the most fundamental factors in residential satisfaction of AIP seniors. Research has been conducted on synergies between architectural design and gerontology, with emphasis on qualitative assessment of architectural quality as well as quantitative evaluation of accessibility and usability in architectural spaces. Lessons are also learned that the successful creation and use of AIP oriented residence require broad stakeholder involvement and collaboration among policy-makers, care planners, designers, constructors and representatives of the senior community (Andersson 2011).

Although no research has explicitly discussed the use BIM for AIP projects yet, several studies (e.g. Han et al 1999, Lopez del Puerto and Clevenger 2010, Yan et al 2011) demonstrated the capacity of BIM in facilitating and visualizing accessibility design in compliance with the Americans with Disability Act (ADA), which is a major criterion employed by AIP projects. Affordability and sustainability are also important principles to be examined when constructing the ideal physical environment for aging in place (Dowhower 2010). In addition to its technical strength, BIM is also adopted by project teams to streamline communication and enhance collaboration among stakeholders. As a new but vetted industry paradigm, BIM is expected to significantly improve the design, visualization and simulation workflow in housing development that is oriented to aging in place. The possible linkage between BIM and Geographic Information System (GIS) is expected to strengthen the bond of individual housing environment with its community context, which is the essential tenet of aging in place.

RESEARCH OBJECTIVE & METHODOLOGY

The overarching goal of the DFA project is to explore the potential and implementing strategies of BIM in AIP efforts. The objective of this paper is to develop a proof of concept through a 360 degree assessment of the feasibility and potential benefits of using BIM in AIP projects. The major tasks to accomplish the assessment include: 1) conducting a series of perception surveys; and 2) constructing a prototype AIP model with common BIM applications.

Perception survey. A total of three (3) surveys were carried out, targeting senior citizens, AIP practitioners and college students as audiences respectively. These surveys used convenience sampling due to the fact that the participants were selected based on their background relevance to this project. The bias associated with convenience sampling can be justified by the fact that the purpose of this pilot study was not to represent the entire population but rather to obtain the basic data and observe the feedback on the issue under investigation from a particular subpopulation. Only descriptive statistics were used in the survey result analysis. The perception surveys were devised to assess and compare opinions on and attitudes towards BIM utilization in AIP projects from stakeholders in the DFA community. The survey findings help identify deficiencies in current AIP practices, the challenges and opportunities for BIM intervention to facilitate the paradigm shift, and the desired BIM capacities and external incentives needed in place to expedite the transition.

Prototype model. A prototype model is created using common BIM applications with embedded AIP criteria, for instance, accessibility design, turnaround design, materials and furnishing specifications. The model serves multiple purposes including design assistance, visualization and communication, compliance checking and evaluation. At the time of this paper, the prototype model is still being developed and has limited contents. Nevertheless, a preliminary evaluation is still performed using the metrics adapted from AIA (2010) to demonstrate the feasibility of BIM implementation in AIP projects.

RESULTS & ANALYSIS

Survey findings. The first survey was conducted through face-to-face interviews with senior citizens, discussing their attitudes towards aging in place, preferred options and perceived challenges that might prevent them from doing so. The second survey with the industry practitioners had a focus on more technical aspects of aging in place, probing for professional's opinions on whether BIM might be a facilitator in AIP projects. The third survey looked at college education as a critical source of intellectual preparedness for future implementation of BIM in AIP projects.

Survey #1: understand the senior population. Twelve (12) local senior people (age 65 or older) participated in this survey via face-to-face interviews. The survey found out the major factors that affected their decision-making on aging in place including *health conditions* (83.3%), *life style* (83.3%) and *costs* (66.7%). The majority (66.7%) of them would prefer to stay in current residence for aging in place, with necessary renovation or remodeling that cost no more than $10,000 (58.3%). When necessary, over half (58.3%) of them would spend no more than $100,000 for new residence purchasing. Overall, 66.7% of them were planning to aging in place.

Housing features that could make the AIP option more viable were discussed during the interview. The participants considered the *functionality*, *accessibility*, *layout/floor plan*, *location/subdivision*, and *overall affordability* to be most desirable features. They were also given a detailed list of design criteria for AIP projects, from which they ranked *accessible bathroom*, *single-story floor plan*, *36" hallway* and *non-slip flooring* as the highest priorities.

The survey then inquired the perceived obstacles that might prevent the seniors from aging in place. A major challenge came from the fact that 11 out of the 12 participants were living in residences with stairs. In order to make the house AIP-friendly, certain remodeling work became inevitable. In this case, these seniors were very concerned about how well architects/contractors understood their unique needs, since the majority of them (75%) could barely read blueprints and did not know how to communicate with architects/contractors through CAD drawings and specifications. So the survey touched upon the possible advantages of using BIM, instead of traditional 2-D based approach, to deliver the intent of AIP design and engage them in the design process. A simple housing model created using Revit was presented to the senior participants, and some interactions were attempted, such as asking them to identify the building components, choose floor plan and layouts, select finishes, and view bathroom details. With such simple first-hand experience with BIM, 11 out of 12 seniors found it much easier to accurately describe and visualize what they really needed in order to age in place.

Survey #2: professionals' perceptions on BIM for AIP. The second survey was deployed online. The email listservs of the Certified Aging-in-Place Specialist (CAPS) from the National Association of Home Builders (NAHB) and the DFA Knowledge Community from the American Institute of Architects (AIA) were used for sampling. A total of 105 professionals responded and 42 completed questionnaires were obtained. The respondents were majorly *Architects* (60%), *Contractors* (13%) and *Interior Designers* (8%). Sixty-one percent (61%) of them had more than 5-year

experience in the field of aging in place. With a BIM definition provided (cited from the National BIM Standard), most respondents (59%) claimed intermediate or advanced knowledge in BIM and its application, while about a quarter (24%) of them had never heard about BIM. In terms of software usage, *AutoCAD* (49%), *Autodesk Revit* (44%) and *Sketchup* (34%) were the top 3 most popular applications. Quite a few respondents (15%) still preferred hand-sketching to any computer applications.

AIP projects were challenging, as most respondents concurred. In particular, they found "*to convey the design intentions efficiently and accurately to clients who don't know how to read blueprints*" (67%) a major test. Equally difficult were "*to understand customer needs and expectations at the desired level of detail*" (56%), "*to experiment and compare possible alternative design options*" (54%) and "*to predict the appearance and performance of the finished project*" (51%). A closer look at these typical challenges in existing AIP practices led to the conclusion that *visualization, communication* and *fast prototyping* are among the key factors in determining the delivery of AIP projects.

A major goal of this survey was to find out professionals' perceptions towards BIM and its potential use in AIP projects. Respondents heard a lot of discussions on the benefits of BIM and its advantages over traditional CAD applications and hand-sketching. The most celebrated benefits that received 50% or more votes included "*dynamic graphic representation and visualization for showcasing and presentation*" (64%), "*robust modeling features with friendly user interface*" (53%), "*improved project productivity and profitability*" (53%), and "*easy generation of blueprints and other project documentation*" (50%).

In regard to implementing BIM in AIP projects, two major aspects were examined: 1) What are the most helpful features of BIM for integrating AIP design criteria; and 2) What are the most desired capacities of BIM for AIP design review for code compliance. According to the survey, professionals considered "*rendering capacities and presentation features*" (58%), "*built-in and expandable library with real manufacturer and vendor data*" (56%), "*modeling functionality with design options*" (53%) and "*user interface with walk-through capacity*" (50%) to be most valuable features for AIP integrated design.

Existing practice of code-compliance review for AIP projects had major deficiencies in maintaining the consistency in code requirements interpretation by different code officials, from one jurisdiction to another. Design review process today is still mostly manual and error-prone, which also tends to be time-consuming and expensive. BIM for automated code-compliance review is still a developing concept but several studies have demonstrated its feasibility, for instance, the Fiatech Autocodes project (see www.fiatech.org for more information). In this survey, participants were asked to make a wish list for desired capacities of BIM for AIP design review for code compliance, as summarized in Figure 1.

It was encouraging to recognize that among the respondents, a considerable percentage (32%) of them had actually done AIP projects using BIM, with another 17% of them had been exposed to conversations with colleagues and other external resources on the potential use of BIM in AIP projects. Respondents quoted "*BIM education*" (62%) and "*positive Return on Investment*" (51%) as major incentives to promoting the use of BIM in AIP projects. They expected to see "*Medium*" level (35%

to 65%) of BIM adoption in the market in 3 to 5 years.
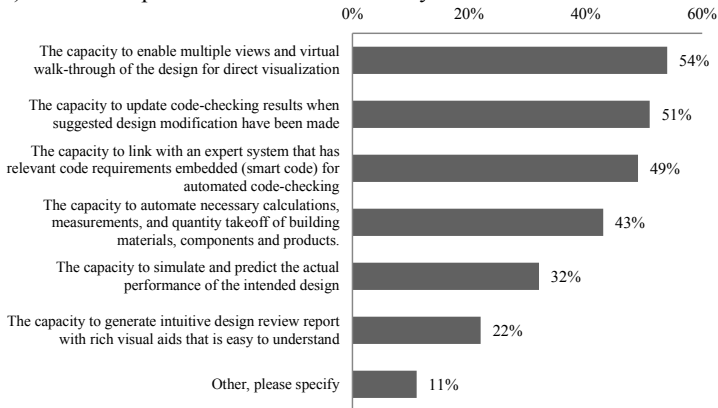
Figure 1. Desired capacities of BIM for AIP design code-compliance review.

Survey #3: BIM in educating future professionals. A short questionnaire was given out in a senior design studio class in the Interior Design program at Georgia Southern University. A total of 19 students responded. They were requested to briefly discuss their learning experience with and application of BIM in course projects. All these students had been taught to use BIM solutions such as Autodesk Revit as well as traditional CAD applications. The majority of them (89.5%) preferred Autodesk Revit to any other available programs when working on course projects. They cited "*save a lot of time*" (63.2%) and "*greatly improve project quality*" (100%) as the major incentives of choosing Revit.

Course projects of the senior design studio class in the Interior Design program usually require understanding of ADA codes, universal design and AIP principles. Students made very positive comments on project outcomes using BIM, since they felt that with BIM it was easy "*to visualize design outcomes and make adjustment*" (100%), "*to achieve more accurate design outcomes in compliance with design criteria and code requirements*" (84.2%), and "*to generate design documents and renderings to present to clients*" (68.4%). Overall, they believed that use of BIM in course projects could result in "*much better learning outcomes*" (63.2%).

Prototype AIP model. The results of the three surveys provided good insights of the benefits and challenges associated with potential use of BIM in AIP projects. To further experiment the idea in a more practical manner, a prototype model was being developed to serve the purposes of: 1) exhibiting an AIP-integrated design workflow; and 2) exploring the AIP criteria compliance check via the virtual environment constructed in BIM. The model embodied the AIP principles in both qualitative and quantitative manner. It is qualitative in terms of the space layout, material & finishing specification and indoor environment, to name a few. It is quantitative because of the spatial allocations and accessibility requirements for each specific room that were taken into account. For instance, doors in open space should be 36" wide with a level floor transition for possible wheelchair rolling over. Doors should also have adequate

push/pull clearances: 15"-18" on pull side of door handle and 12" on push side of door handle. All lighting switches and outlets are mounted 24" above finished floor (AFF) and door handles 36" AFF according to accessibility standards. All aisle ways from furnishings to adjacent structures must be at least 36" and hallways are also 36" or greater, to name a few. The prototype model is still under developing at the time of this paper. Nevertheless, some initial design details are exhibited in Figures 2a-2d. Local code officials were also invited to evaluate the prototype model for its validity and usefulness in exemplifying value of BIM in AIP projects.
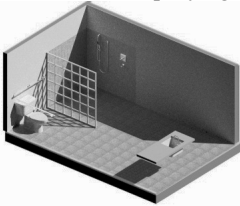
2a. Bathroom.
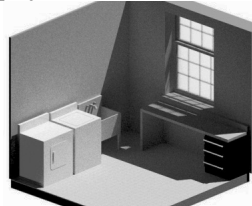
2c. Laundry room.
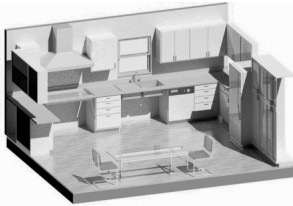
2b. Open kitchen.
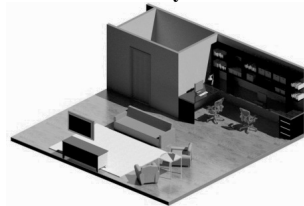
2d. Open living room.

Figure 1a-1d. Snapshots of prototype model details.

CONCLUSIONS & FUTURE WORK

This paper presented the initial results of a project that investigates the feasibility and strategies of implementing BIM in AIP projects. Through a series of perception surveys, a 360 degree assessment was conducted. The results of the survey indicated that senior people would prefer to age in place when their health conditions and the housing environment allow so. A major barrier for the AIP efforts is the lack of means in existing practices for efficient communication and understanding between the professionals and the senior population. BIM has been valued as an advancing modeling technology as well as a better platform for client-professional collaboration. The undergoing efforts of developing a fully functional AIP prototype model are anticipated to offer the stakeholders a new business paradigm that can revolutionize current best practices. Future work will focus on completing and validating the prototype model, and also look for opportunities to utilize the model in real AIP project delivery.

REFERENCES

American Institute of Architects (AIA). (2010). *AIA Design for Aging Post-Occupancy Evaluation: Evaluator's Toolkit 2010.* AIA, Washington, D.C.
Andersson, J.E. (2011). "Architecture for the silver generation: Exploring the meaning of appropriate space for ageing in a Swedish municipality." *Health &*

Place, 17(2), 572-587.

Ball, M. S. (2004). "Aging in Place: A toolkit for local government." *Community Housing Resource Center*, <http://www.aarp.org/content/dam/aarp/livable-communities/plan/planning/aging-in-place-a-toolkit-for-local-governments-aarp.pdf> (Nov. 5, 2012)

Cutchin, M.P. (2003). "The process of mediated aging-in-place: a theoretically and empirically based model." *Social Science & Medicine*, 57(6), 1077-1090.

Cutler, L.J. (2007). "Physical environments of assisted living: research needs and challenges." *The Gerontologist*, 47(3), 68–82.

Devlin, A.S., and Arneill, A.B. (2003). "Health care environments and patient outcomes. A review of literature". *Environment and Behavior*, 35(5), 665–694.

Dowhower, J.F. (2010). *Adapting Building Information Modeling for Affordable & Sustainable Housing*, Master Thesis, University of Texas at Austin.

Farber, N., Shinkle, D., Lynott, J., Fox-Grage, W., and Harrell, R. (2011). "Aging in Place: A State Survey of Livability Policies and Practices." <http://www.ncsl.org/documents/transportation/aging-in-place-2011.pdf> (Nov. 5, 2012)

Greenfield, E.A., Scharlach, A., Lehning, A.J., and Davitt, J.K. (2012). "A conceptual framework for examining the promise of the NORC program and Village models to promote aging in place." *Journal of Aging Studies*, 26(3), 273-284.

Han, C., Kunz, J. and Law, K. (1999). "Building design services in a distributed architecture", *Journal of Computing in Civil Engineering*, 13(1), 12-22.

Lopez del Puerto, C., and Clevenger, C. (2010). "Enhancing Safety throughout Construction using BIM/VDC." *Proceedings of the 2010 BIM-Related Academic Workshop,* Ecobuild, Washington D.C., Dec.6-10.

McGraw-Hill Construction. (2008). *Building Information Modeling: Transforming Design and Construction to Achieve Greater Industry Productivity,* The McGraw-Hill companies, New York.

McGraw-Hill Construction. (2009). *The Business Value of BIM: Getting Building Information Modeling to the Bottom Line,* The McGraw-Hill companies, New York.

Morley, J.E. (2012). "Aging in Place (Editorial)."*Journal of the American Medical Directors Association*, 13(6), 489-492.

Olshansky, S.J., Goldman, D.P., Zheng, Y., and Rowe, J.W. (2009). "Aging in America in the Twenty-First Century: Demographic Forecasts from the MacArthur Foundation Research Network on an Aging Society." *The Milbank Quarterly,* 87(4), 842-862.

Rioux, L., and Werner, C. (2011). "Residential satisfaction among aging people living in place." *Journal of Environmental Psychology*, 31(2), 158-169.

U.S. Census Bureau. (2012). "2012 National Population Projections, Table 2. Projections of the Population by Selected Age Groups and Sex: 2012-2060." < http://www.census.gov/population/projections/files/summary/NP2012-T2.xls> (Dec. 30, 2012)

Yan, W., Culp, C., and Graf, R. (2011). "Integrating BIM and gaming for real-time interactive architectural visualization." *Automation in Construction*, 20(4), 446-458.

Synthesis of Existing BIM Maturity Toolsets to Evaluate Building Owners

Brittany Giel[1] and Raja R.A. Issa[2]

[1] Rinker School of Building Construction, Univ. of Florida, P.O. Box 115703, Gainesville, FL 32611-5703; PH (352) 339-0237; FAX (352) 846-2772; Email: b1357g@ufl.edu
[2] Rinker School of Building Construction, Univ. of Florida, P.O. Box 115703, Gainesville, FL 32611-5703; PH (352) 273-1152; FAX (352) 846-2772; Email: raymond-issa@ufl.edu

ABSTRACT

As designers and contractors become more advanced users, facility owners have also begun to make significant strides toward adopting Building Information Modeling (BIM) within their organizations. In fact, many larger owner organizations are now contractually requiring BIM on their construction projects. However, operationalizing the technology during the Operations and Maintenance (O&M) phase is a consistent hurdle that many owners have been unable to overcome thus far. The improvement of building owners' BIM competency is critical to reaping the full benefits of BIM over the building life cycle. Therefore, this research aims to identify what key characteristics building owner organizations must possess to become mature BIM users. An in depth analysis of the existing BIM maturity assessment tools was conducted to make comparisons between the different variables measured. Suggestions are made for adding variables that have not yet been considered in the evaluation of building owner BIM competency and which may also be important.

INTRODUCTION

Facility owners may play the most significant role in the Architecture, Engineering, Construction and Operations (AECO) industry's quest to achieve life cycle use of Building Information Modeling (BIM). Yet, their deployment of BIM is still infrequent at best. Some owner organizations have begun to require BIM during design and construction, but lack the technical and organizational knowledge required to operationalize a model or its data post construction. It is imperative that the BIM competencies of facility owners be evaluated in order to encourage the continued use of BIM over a building's life cycle.

This paper presents a detailed comparison of the different BIM maturity assessment tools currently being utilized and seeks to determine whether the existing toolset adequately addresses the specific needs and information requirements of facility owners. It aims to identify what critical factors must be measured in the evaluation of building owners' BIM competency. Therefore, the research questions posed were aimed at determining the factors that are important in rating the BIM maturity of all AECO organizations implementing BIM and at determining which factors differentiate owner organizations from other stakeholder groups that are implementing BIM.

BACKGROUND

A *Maturity Model* may be defined as a structured group of elements and practices which characterizes effective processes and/or products usually consisting of a limited number of maturity levels which are sequential and characterized by specific requirements at each level (Khoshgoftar and Osman 2009). Rooted in the quality management and improvement philosophies of Crosby, Deming and Juran and the Software Engineering Institute's (SEI) Capability Maturity Model (CMM) for software process improvement, the concept has been adapted in numerous domains, but particularly in the IT sector (Succar 2009). Paulk et al. (1993) first distinguished *Process Maturity* to describe the extent to which a specific practice is explicitly defined, managed, measured, controlled and considered effective. He theorized that organizations utilizing repeatable defined processes that are frequently updated may produce more mature results.

Since the 2007 publication of the NBIMs v1, many researchers have applied these same process improvement ideologies to the BIM domain. In fact, there have been many adaptations of BIM maturity assessment tools developed in recent years. Most of these tools have followed the general structure of the SEI's original CMM and have been adapted to the building industry and BIM (NIBS 2007, Succar 2009, CIC Research Group 2012). However, others have approached evaluation from a quantitative benchmarking perspective (Sebastain and van Berlo 2010, CIFE 2009).

NBIMS' Capability Maturity Model (CMM). The first attempt at evaluating BIM maturity was proposed in 2007 in Version 1 of the National BIM Standard. The BIM Capability Maturity Model (CMM) was developed to define the quality and quantity of information required to be considered a "minimum BIM" (NIBS 2007). The tool evaluates BIM-assisted projects based on 11 areas of interest against 10 increasing levels of maturity. Certification levels possible in the CMM include: minimum BIM, certified, silver, gold and platinum, for a maximum score of 100 points. As of 2011, the minimum score required for the distinction of minimum BIM was set to 60 points (NIBS 2012).

BIM Maturity Index (BIMMI). In response to some of the NBIMs' CMM limitations, Succar (2009) developed his own BIM Maturity Index (BIMMI). The framework assesses the quality of teams and organizations rather than evaluating information management on a BIM-assisted project. It is scalable for different organization types and utilizes five maturity levels based on 12 Key Maturity Areas (KMAs). Unlike the CMM which was created for internal self-assessment, his model offers four different types of evaluation, many of which utilize the rating of an external auditing agency.

VDC/BIM scorecard. During the same period, Stanford's Center for Integrated Facility Engineering (CIFE) began developing an industry performance rating system called the BIM scorecard to help benchmark Virtual Design and Construction (VDC) practices. The scorecard uses the results of four input survey forms related to the areas of Planning, Adoption, Technology and Performance to measure the degree of innovation being achieved on BIM-assisted projects. Instead of maturity levels, each

dimensional measure is scored on a percentile system with different percent ranges addressing: Conventional Practice, Typical Practice, Advanced Practice, Best Practice, and Innovative Practice. The tool's intent is to assess and compare the maturity of VDC processes against industry standards and to make recommendations to organizations on how to improve their BIM performance (CIFE 2012).

BIM QuickScan Tool. A tool similar to CIFE's scorecard program was introduced in the Netherlands in 2009 by the Netherlands Organization for Applied Scientific Research - TNO. TNO's BIM QuickScan tool is a benchmarking instrument to assess the BIM performance of firms executing technology and processes. It rates organizations based on four distinct chapters of criteria: organization and management, mentality and culture, information structure and flow, and tools and applications. Each chapter is comprised of weighted key performance indicators (KPIs) that are collected from responses to multiple choice questions collected by a BIM consultant based on observation and interviews with BIM management personnel. Questions are also categorized into ten specific aspects including: strategic, organization, resources, partners, mentality, culture, education, information flow, open standards and tools (van Berlo et al. 2012). Similar to the other assessment tools, the scores for each category are mapped into a radar diagram to help organizations visualize areas for improvement. Recommendations are made in the form of a report by the consultant to aid organizations in the continuous improvement of BIM processes (Sebastian and van Berlo 2010).

BIM Maturity of Building Owners. Most of the available assessment tools have been intended for stakeholders to evaluate themselves and where they stand in terms of BIM implementation. However, these models seldom address the maturity of building owner organizations. Only two distinct models thus far have suggested the perspective of owners to be unique. The only published tool designed *by* an owner organization thus far has been Indiana University's (IU) *BIM Proficiency Matrix*. The tool is a standard means for evaluating the BIM experience of potential designers and contractors on new projects. The matrix is used to score applicants based on eight general categories including: physical accuracy of the model, the presence of an Integrated Project Delivery (IPD) methodology, calculation mentality, location awareness, content creation, construction data, as-built modeling and FM data richness. Potential stakeholders must provide a description and concrete example of past projects that they have participated in which addressed each BIM proficiency category in an MS Excel template (Indiana University 2010).

More recently, the CIC Research Program's (2012) Facility Owner's Guide to BIM Execution provided a template matrix for owners that are beginning to require BIM in order to aid them in the assessment of their organizational BIM execution efforts. The matrix is divided into six categories based on what were deemed to be key BIM *Planning Elements* by the CIC Research Program including: *Strategy, BIM Uses, Process, Information, Infrastructure* and *Personnel.* Owners are encouraged to rate their organizational BIM maturity from 1-5 in each of the sub categories of BIM execution planning as a means to encourage continuous improvement.

SYNTHESIS OF EXISTING BIM MATURITY FACTORS

To summarize the similarities and differences between the leading BIM assessment tools, a comparative matrix was developed. It outlines the different variables which have been incorporated in each tool and where potential overlap occurs. The variables have been categorized into six major areas including: Planning Competencies, Technical Competencies, Personnel Competencies, Managerial Competencies, Process Competencies, and BIM Requirements. All of these are seen as potential areas in which building owners' BIM competency may need to be evaluated. These variables are also further categorized into process-oriented variables and product-oriented variables based on the evaluation intent of the models.

Figure 1 describes the variables which may be categorized as Planning Competencies. As shown, there is a greater emphasis placed on organizational process variables than variables related to the model or project (product) in the existing maturity evaluation toolset. The bold dot in the matrix represents the inclusion of that specific variable in the corresponding assessment tool shown. Planning competency variables include factors related to the strategic planning efforts of an organization. They describe the ability of an organization to a construct a course of action for BIM procedures internally through objectives, standards, and guidelines documentation as well as a quality assurance plan for how documentation will be revised over time.

1. PLANNING COMPETENCIES							
Variable Type	Maturity Variables	Model					
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
Process	Presence of Organizational Mission and Goals						●
Process	Leadership: BIM vision / Planning-Objective: Objectives Documentation, Management objectives, outline for which stakeholders benefit from objectives/ Vision and Strategy /Presence of BIM Vision and Objectives		●		●	●	●
Process	Planning-Preparation: VDC budget /Financial resources and partnership at project level				●	●	
Process	Regulatory: BIM Guidelines and standards documentation/Planning-Standard: VDC Guidelines, VDC Guides, Contribution to VDC guideline development		●		●		
Process	Creation of a BIM Execution Plan/Planning-Standard: Contents covered in BEP			●	●		
Process	Adoption-Organization: Frequency of process improvement efforts/ quality assurance				●	●	

Figure 1. Synthesis of Planning Competency Variables

Figure 2 shows the variables in the existing BIM maturity toolsets which may be considered Technical Competencies. Variables in this competency area include factors related to the technical changes that must be made within an organization to implement BIM. This includes changes to software, hardware and networking to prepare for the use and storage of BIM deliverables. As shown, there is consensus among the existing BIM maturity tools regarding the significance of technology in the evaluation of BIM processes.

Figure 3 shows the variables considered to be part of Personnel Competencies. These include factors related to the knowledge, skills, and attitude of the people within an organization who are employed to facilitate BIM. As shown, a major factor which the majority of the existing BIM Maturity toolsets seem to consider is whether the BIM roles and responsibilities are clearly defined for internal

employees as well as project team members. Additionally, the presence of education and training was a common factor considered in several of the BIM Maturity toolsets.

2. TECHNICAL COMPETENCIES							
Variable Type	Maturity Variables	Model					
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
Process	Software changes/ Planning-Preparation: # of VDC software and Technology-Integration: Software adequacy/Type of software and BIM toos used/ Software		●		●	●	●
Process	Hardware/ Technology-Integration: Hardware Adequacy/Hardware		●		●		●
Process	Network Changes/ Planning-Preparation: Data Sharing Method used/Use of Model server and type and capacity of server		●		●	●	
Process	Physical space						●

Figure 2. Synthesis of Technical Competency Variables

3. PERSONNEL COMPETENCIES							
Variable Type	Maturity Variables	Model					
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
Process	Adoption-Organization: # of stakeholders with VDC resp, Ratio of stakeholders creating VDC content, Ratio of FTE involved with VDC/ Organization Structure/ Organizational Hierarchy				●	●	●
Process	Roles or Disciplines/ Human Resources / Model Managers Role defined/ Distribution of roles and tasks/ Roles and Responsibilities / Interdisciplinary integration and responsibilites defined on BIM projects	●	●	●			●
Process	Leadership/Adoption-Organization: # of stakeholders motivated to leverage VDC, # of stakeholders involved in VDC decision making, Direct and potential benefactors of VDC/ Management Support		●		●		●
Process	Presence and influence of BIM coordinator/ Presence of BIM Champion				●		●
Process	Creation of BIM planning committee						●
Process	Human Resources' competencies/Adoption-Organization: VDC Experience of Organization and Lead Personnel, VDC skill level of project team members		●		●		
Process	Change readiness of staff						●
Process	BIM acceptance among staff and workers and group and individual motivation					●	
Process	Preparatory: education and research/Education		●				●
Process	Preparatory: training/Adoption-Organziation: Ratio of Stakeholders trained in VDC, Scale of VDC Training/Frequency of VDC training, % of time staff spends working with applications/ Knowledge management and training/ Training		●		●	●	●

Figure 3. Synthesis of Personnel Competency Variables

Figure 4 shows the variables considered to be Managerial Competencies. Variables in this competency area include factors related to the policies and decisions made to manage BIM-assisted projects. The factors listed are common to all stakeholder organizations implementing BIM, but the policies that must be implemented in requiring BIM over the building life cycle are perceived to be an area where owner organizations differ greatly from other stakeholder categories.

Figure 5 shows the variables which may be categorized as Process Competencies. Variables in this competency area include factors related to how BIM

is used both internal and external to the organization. This would coincide with different use-cases and applications of BIM in each lifecycle phase. Though this may vary from project to project, the capability of an organization to implement and/or require that the model be used in these ways is perceived as a key indicator of owner maturity.

Variable Type	Maturity Variables	Model					
4. MANAGERIAL COMPETENCIES							
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
Process	Infrastructure: How BIM knowledge is shared and retained within an organization/Knowledge management & knowledge and skills		●			●	
Process	Contractual changes		●				
Process	Delivery Method/IPD Methodology/ Adoption-Process: Proj Delivery Method	●		●	●		
Process	Regulatory: model standards/Planning-Standard: Naming conventions for objects/object libraries		●		●	●	
process	Regulatory: model protocols / establishment of model view definitions and supporting rules		●			●	
Process	Adoption -Process: Frequency of Project wide mtgs (i-rooms), Response to RFIs				●		
Process	Adoption-Organization: Stakeholder Involvement avg # of phases covered				●		
Process	Risk Management strategies		●				
Process	Regulatory: specifically Benchmarking and R&D strategies/Performance-Quantitative: VDC mgmt objectives tracked, assessment of objectives vs actual performance, alignment btw objectives in planning areas and technology area, % of RFI on time vs expected, Unforsee change orders as % of total cost, field inititated change order as % of total cost		●		●		
Process	Regulatory: specifically Benchmarking and R&D strategies/ Performance-Qualitative: Model uses assessment in technology, meeting satisfaction level x meeting importance level, diamond of user emotions		●		●		

Figure 4. Synthesis of Managerial Competency Variables

Variable Type	Maturity Variables	Model					
5. PROCESS COMPETENCIES							
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
Process	Business Processes/Organizational Processes mapped	●					●
Process	Business Processes/ Project Processes mapped	●					●
Process	BIM Capability Stage		●				
Process/Model	Life Cycle Views/Adoption-Process: Avg # of VDC application of phases covered and Technology-Integration: % of stakeholders using model, % of team members using model, % of models for communication with client	●			●		
Process/Model	Addressed by entire matrix /Technology-Maturity: Model uses/ use of modeling/ Project Uses of model			●	●	●	●
Process/Model	Operational Uses of model						●

Figure 5. Synthesis of Process Competency Variables

Finally, Figure 6 shows mainly the product-oriented model variables which comprise the BIM Requirements area. Variables in this competency area include established requirements for BIM deliverables and are further broken down into *Information Requirements* and *Geometric Requirements*. The variables which fall into this category are mainly associated with aspects of a final product versus the

organization implementing the product. Therefore, many of these variables may be valued higher in the evaluation of building owner BIM competency. As indicated, the incorporation of Open BIM standards and how interoperability is addressed is one factor which was considered by many of the existing tools.

Variable Type	Maturity Variables	6. BIM REQUIREMENTS					
		Model					
		2007 BSA:ICMM	2009 Succar: BIMMI	2009 IU: BIM Proficiency Matrix	2009/2011 CIFE: VDC Score card	2009/2010 TNO: BIM QuickScan	2012 Penn State: Owner Maturity
	Information Requirements						
Process/Model	Products and Services: model progression specification/ Model element breakdown structure		•				•
Process/Model	Products and Services: model progression specification/ Technology-Integration: Average LOD, LOD Adequacy/LOD		•		•		•
Process/Model	Facility Data Needs established						•
Process/Model	Data Richness/Type of data exchange and data in each phase	•				•	
Process/Model	Internal and external information flow					•	
Process/Model	Interoperability & IFC Support/model information export capability/ Technology-Integration: Degree of Model elements exchanged and Information loss after model exchange/Open ICT Standards	•		•	•	•	
model	Information Accuracy	•					
model	Calculation mentality			•			
model	Content Creation: Design Intent			•			
model	Construction Data: Quantity Takeoffs			•			
model	Construction Data: Object Scheduling			•			
model	Construction Data: Material Procurement			•			
model	FM Data Richness: Space Management			•			
model	FM Data Richness: Asset Management			•			
model	Content Creation: Manufacturer's Specific/ Technology-Coverage: % OF 3D modeled product elements				•	•	
	Geometric Requirements						
model	Graphical Information/ Geometrically correct content	•		•			
model	Design Requirements			•			
model	Design side collision detection			•			
model	Spatial capability/ location awareness	•		•			
model	existing enviroment integration			•			
model	global accuracy			•			
model	post bid model documentation accuracy			•			
model	coordination modeling accuracy			•			
model	As-built Record model accuracy			•			
	Other						
Process/Model	Change Management	•					
Process/Model	Timeliness/Response	•					

Figure 6. Synthesis of BIM Requirements Variables

CONCLUSIONS

Analysis of the existing BIM maturity tools revealed many similarities among the different models. First, most of the tools tend to categorize assessment variables. Whether called *Chapters, Areas, or Categories,* their intended purpose is to group and present assessment factors based on their logical similarities. Another common trend is the use of weighting systems to allocate points to the final assessment score/rating. For example, greater weight is placed on areas of interest like Interoperability/IFC Support and Information Accuracy in the NBIMs' CMM; whereas the areas of Technology and Performance are weighted higher in CIFE's BIM Scorecard program. The empirical reasoning for many of these weightings remains unclear to the author, as few to none of these tools have been objectively validated (Dib et al. 2012).

The defining difference between the existing tools is their evaluation approach. The NBIMs' CMM and the BIM Scorecard assess an individual building project assisted by BIM, focusing more on the comprehensive appraisal of a product. But more commonly, tools such as the BIMMI, BIM QuickScan, Penn State's Owner maturity model, and the IU BIM proficiency matrix assess the maturity of organizations implementing BIM processes. Thus, one of the greatest weaknesses of the existing tools is their inability to measure the BIM maturity of both the model and

the organization simultaneously (Sebastian and van Berlo 2010). Additionally, only two of the existing models have specifically addressed BIM maturity from the perspective of building owners (CIC Research Program 2012, Indiana University 2009).

Out of the BIM maturity tools that were analyzed, there were two variables identified which may have significant impact on the assessment of building owners' BIM competency, but which have not been explored in the existing tools. On such variable is the selection process used to qualify designers and contractors on BIM-assisted projects. This would refer to whether and how the BIM experience of each stakeholder is considered before the project team is selected and gets awarded the contract. It may be one of the most important factors contributing to the success of a BIM-assisted project at the onset. Additionally, another variable that has not been explored by other researchers is whether an owner organization has a plan of action in place for how the model will be updated as renovation/re-modeling activities take place. This would be a document internal to owner organizations which outlines protocols for updating building information over a building's life cycle.

REFERENCES

CIFE (2011). "CIFE VDC Scorecard and BIM Scorecard." Retrieved from: <http://cifevdc.weebly.com>

Computer Integrated Construction (CIC) Research Program (2012). "BIM Planning Guide for Facility Owners," Pennsylvania State University: University Park, PA. Retrieved from: <http://bim.psu.edu>

Dib, H., Chen, Y. and Cox, R. (2012). " A Framework for Measuring Building Information Modeling Maturity Based on Perception of Practitioners and Academics Outside the USA," *Proceedings 29*th International Conference *CIB W78 2012,* October 17-19. Beirut, Lebanon, paper 34.

Indiana University (2009). "Building Information Modeling (BIM) Guidelines and Standards for Architects, Engineers, and Contractors." Retrieved < http://www.indiana.edu/~uao/iubim.html>

Khoshgoftar, M. and Osman, O. (2009). "Comparison of Maturity Models," Proceedings from the 2nd IEEE International Conference on Computer Science and Technology, Aug. 8-11, 297-301.

National Institute of Building Sciences (NIBS) (2007). National Building Information Modeling Standard - Version 1.0 - Part 1: Overview, principles and Methodologies: NIBS.

National Institute of Building (NIBS) (2012). National Building Information Modeling Standard - Version 2.0 –Chapter 5.2 Minimum BIM: NIBS.

Paulk, M., Curtis, B. and Chrissis, M.B., and Weber, C. (1993). "Capability Maturity Model for Software, V. 1.1," Technical Report CMU/SEI -90-TR-24, Software Eng. Institute.

Sebastian, R. and van Berlo, L. (2010). "Tool for Benchmarking BIM Performance of Design, Engineering and Construction Firms in the Netherlands," Architectural Engineering and Design Management, 6, 254-263.

Succar, B. (2009). "Building Information Modeling Maturity Matrix." *Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies*, Information Science Publishing, Hershey, PA.

Van Berlo, L., Dikkmans, T., Hendriks, H., Spekkink, D., and Pel, W. (2012). "BIM QuickScan: Benchmark of Performance in the Netherlands," *Proceedings 2*9th International Conference *CIB W78 2012,* October 17-19. Beirut, Lebanon, paper 30.

Using model updating to predict the failure of reinforced concrete elements

Yaqub M. Y. Rafiq[1] and M. Al-Farttoosi[2]

[1]Associate Professor University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK.
mrafiq@plymouth.ac.uk
[2]Researcher University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK.
mahdi.al.farttoosi@postgrad.plymouth.ac.uk

ABSTRACT

Finding an accurate analytical model that is a good match for experimental data is not a straightforward task. Generally analytical model parameters are tuned, using previously tested methods, to achieve an acceptable match for the experimental results. Model updating is a popular method that has been widely used since the early 1990s for this purpose. In this paper, research carried out for predicting the behaviour of masonry wall panels has been extended to find a suitable non-linear model for reinforced concrete (RC) beams that matches the experimental laboratory results. Data from laboratory experiments carried out by other researchers are used to confirm the validity of this analytical model. A Genetic Algorithm (GA) tool is used for the model updating process. The GA optimisation process finds a set of optimum parameters for use in the FE model. The outcome of this study was a close match with the experimental results, thus utilisation of this technique will dramatically reduce the number of expensive laboratory experimental tests which would otherwise be necessary.

KYEWORD: Model updating, System Identification, Optimisation, Genetic Algorithms,

INTRODUCTION

Model updating techniques have shown to be an effective way of verifying the suitability of analytical models. This is a reverse problem solving process, which will naturally create a cluster of good solutions around the optimum point that can all produce desirable results. Due to the interaction and compensatory effect of many variables, it would be difficult to find a unique solution using stochastic search techniques such as the GA. Sometimes model updating techniques may result in a model that can reasonably simulate the experimental results. In this paper the genetic algorithm was used to generate alternative analytical models that are a good representation of the experimental results.

According to Friswell and Mottershead (1995), finite elements model updating emerged in the 1990s as a subject of immense importance to design, construction and maintenance of mechanical systems and civil structures. Since its emergence, model updating techniques have been widely used for structural health monitoring, system identification and maintenance of many complex structural systems. Amongst others, bridge structures have been the subject of a number of studies. Brownjohn et al. (2001) studied the sensitivity-analysis-based Finite Element (FE) model updating method and applied this to the condition assessment of bridges. The studies of Brownjohn et al. were inspired by Aktan et al (1996) who documented that about 125,000 of the 585,000 bridges in the USA are deemed deficient. Aktan et al. (1996-1998) conducted a large amount of research on the condition assessment of bridges. According to Brownjohn (2001), FE modelling has been successfully applied in mechanical and aerospace disciplines. This model updating technology has proved to be challenging to apply in civil engineering structures, because of the difficulties in prototype testing and experimental data analysis resulting from the nature, size, location and usage of these structures.

Model updating of bridges generally consists of studying the dynamic responses of bridges to different loading types. Brownjohn and Xia (1999, 2000) outline the model updating procedure to include three main aspects: (1) Selection of 'responses' as reference data which are

normally the measured data; (2) Selection of parameters to be updated, to which changes in the selected responses should be sufficiently sensitive and must be uncertain; and (3) Model tuning which is an iterative process to modify the selected parameters of the structure.

Brownjohn et al. (2010) carried out an ambient vibration survey of the Humber Bridge in the UK. Their work included identification of the equipment and procedures used for the test, comparison of the operational modal analysis (OMA) technology used for system identification and they presented modal parameters for key vibration modes of the complete structure.

Raphael and Smith (1998) used stochastic search as a hybrid reasoning system for complex diagnostic tasks in structural engineering. They proposed that the initial model set is created through model composition and that parameter values determined by stochastic search can lead to more accurate predictions of behaviour. Goulet et al. (2012) presented a study of quantifying modelling uncertainties using the Grand-Mere Bridge located in Canada. Their case study demonstrated that modelling errors involve systematic biases that can affect data interpretation and ultimately the quality of infrastructure-management support.

Goulet et al. (2010) applied multi-model structural performance monitoring that explicitly incorporated uncertainties and modelling assumptions. This approach was applied to measurements from structural performance monitoring of the Langensand Bridge in Lucerne, Switzerland. They demonstrated that measurements from load tests may lead to numerical models that better reflect structural behaviour. This approach sampled thousands of models, starting with general parameterised FE models. The population of selected candidate models was used to understand and predict behaviour, which leads to better structural management decision making.

Numerical model updating techniques were used to predict the behaviour of masonry panels by Zhou (2002) and Rafiq et al. (2003). These authors introduced the concept of stiffness/strength corrector factors which assigned different values of flexural rigidity or tensile strength to various zones within a wall panel. Zhou et al. (2003, 2006) developed methodologies for zone similarity techniques, which introduced an automatic technique assigning corrector factors from a single "base panel". This was tested in the laboratory on any "new panel" for which no laboratory data exists, by comparing similar zones between two panels. These authors used a cellular automata (CA) model that propagated the effect of panel boundaries to zones inside the panel. Further study of the corrector factors by Sui (2007), using model updating techniques, revealed that boundary effects had a major influence on the behaviour of masonry panels. This author demonstrated that inaccurate modelling of boundary conditions of masonry panels leads to a major error in the analytical results. Sui et al. (2008) introduced a numerical method of using boundary parameters to improve the finite element analysis (FEA) predictions of the response of masonry wall panels to lateral loads.

The paper presented by the authors is the initial stage of a PhD programme that studies the behaviour of reinforced concrete beams externally strengthened with carbon fibre reinforced polymer (CFRP) subjected to the impact loading. The research will include laboratory impact testing to induce controlled damage to several RC beams and then these beams are strengthened by carbon fibre reinforced polymer (CFRP) and their resistance and suitability as structural members are evaluated. To decide on a reasonable number of experimental samples, a selective number of RC beam data obtained from laboratory tests in published literature were selected. This data mainly consists of the static load deflection information measured in the labs until the failure of a beam occurs. A Genetic Algorithm (GA) optimisation process was used that varies some critical model parameters within the RC beam. These modified model parameters were then passed to a non-linear FE model. Load deflection curves and ultimate failure load obtained from the FE model is then compared with those obtained in the laboratory.

METHOD

FEA Model Description: In this study, three RC beams, tested by Esfahani et al. (2007), were used. At first, load deflection data from un-strengthened beam test results were used for FEA

validation purposes. The LUSAS finite element commercial package was used to conduct a non-linear analysis for all RC beams. In the modelling process, concrete was modelled as an 8 nodes 2-dimensional Plane Stress Continuum Element (QPM8). A nonlinear (Elastic: Isotropic, Plastic: Multi-Crack Concrete) Model 94 was used to model the concrete. The multi-crack concrete model is a plastic-damage-contact model in which damage planes form according to a principal stress criterion and then develop as embedded rough contact planes (LUSAS Manuals). Steel reinforcement bars were modelled as Non-linear 2D-Bar elements with 3 nodes, using the Stress Potential von-Mises model that considers yield stress with hardening curve gradient properties. Load increments were used in the non-linear analysis process. Analysis was continued until the failure load was reached and load deflection graphs from the FEA results were generated for comparison with the experimental load deflection curves.

Uncertainties related to reinforced concrete elastic and non-linear material properties are very difficult to quantify. This is due to the nature of concrete that depends on many variables (e.g variation in aggregate properties from batch to batch and even within each batch; variation in cement content; water-cement ratio etc.). This was an obvious choice to use as a varying parameter in the analysis and as subject of the optimisation process.

To make the FE process more efficient, a parametric study was conducted to optimise the number and type of elements used in the FE model, simplified support conditions and loading.

In this study a FE model was developed; material properties for each zone (13 in total) were generated by the GA and these were subsequently assigned to every element in the beam. The FE software was then called from within the optimisation algorithm for fitness evaluation purposes. Details of the GA optimisation process are discussed in the following section.

The Genetic Algorithm (GA) Model Description: The purpose of using the GA optimisation process is to find a set of suitable parameters for the RC beam for use in the FE analysis. The outcome of the optimisation process is to find a set of best solutions that is a close fit to the experiment results. In this study, concrete elastic modulus for each zone within the beam was selected to be the main variable of the GA. It should be noted that there is less uncertainty in the value steel elastic modulus and it can be determined relative easily and more confidently from the laboratory test results. Therefore this parameter was not considered as a variable parameter in the optimisation process. The process of optimisation in this particular context is summarised as:

1. The beam in the experimental model was a simply supported beam and it was tested under a standard four point load test. Values of load increments and vertical displacements at various locations along the length of the beam were recorded in the laboratory experiment.
2. For the FE model, a symmetrical half of the beam was modelled. The symmetrical half of the beam was divided into 13 zones as shown in Fig 1. In this model, the zone near the support was finely meshed.
3. Each zone was assigned a different elastic modulus value; thus 13 variables were used in the GA process. A binary string was selected in which each variable was allocated 3 bits. The range of elastics modulus was varied between 0.5 to 2.0 times the original elastic modulus. It is worthwhile to mention that deriving a correct value for the concrete elastic modulus is not precisely defined in Codes of Practice or in published literature. Hence, information from the load deflection curve and also code rules were used to derive the initial elastic modulus value. The main object of the GA optimisation process was to find a set of solutions that result in a good fit for the experiment load deflection curve. The GA process can be briefly describes as follows:
   a. 13 variables used in the GA binary string each of which consisting of 3 bits. These variables were decoded and mapped as:
      map_obj_Value =(1.0/70.0)*[35.0, 50.0, 65.0, 80.0, 95.0,110.0,125.0,140.0]

Resulting in [0.50, 0.714, 0.929, 1.143, 1.357, 1.571, 1.786, 2.0] factors for each variable. The original module (a single value) for the beam elastic modulus was then multiplied by these factors to get the elastic modulus for each zone.
   b. A routine was written to automatically modify the FEA data file for the new set of elastic modulus values.
   c. A non-linear FEA analysis is then conducted for each member of the GA population and the process was continued until the termination conditions were met.
   d. The whole population of the GA was then saved for further scrutiny.
4. The following parameters were optimised in the objective function of the GA.
   a. Value of the failure load to be close to the experimental failure load.
   b. Gradient of the load deflection curve to be close to the experimental gradient.
   c. The overall deflected shape of the beam to be close to the experimental results.
   d. The value of the maximum deflection at the centre of the beam should be close to that of the experimental results.
   e. A weighted sum of the above four objectives. In this process, various weightings were experimented in which knowledge of previous work in this area and a trial and error process was used.
5. At the end of the optimisation process, load deflection curves for a set of suitable candidate solutions [top ranked by the GA solutions] were plotted and compared with the experimental results for the final selection of the elastic modulus values.
6. Once this process was completed, the selected model(s) from the GA were imported to the FE software for further scrutiny.
   A population size of 100 for 50 generations was considered in the GA process. This includes 5000 FEA calls. Each run of the FEA took about 1 minute and the whole GA run took about 3 and half days to complete.

RESULTS AND DISCUSSION

Esfahani et al. (2007) investigated the flexural behaviour of RC beams strengthened with CFRP under static loading. Twelve RC beams were tested using four point loads up to failure. The dimensions of tested beams were 150 mm width, 200 mm depth and 2000 mm length. Fig 2 shows beam dimensions and details. Three of the RC beams were considered as reference beams, not strengthened with CFRP. Whereas the other nine beams were externally strengthened with CFRP and different reinforcement ratios were used. Beams B1-12D-0L, B2-16D-0L and B9-20D-0L were the control beams without strengthening. Different longitudinal bottom bars reinforcement were used, 12mm bar in beam B1-12D-0L, 16 mm bar in beam B2-16D-0L and 20 mm bar in beam B9-20D-0L. In this study, these three beams were used for validation of the GA optimisation process. A comparison between experimental results and finite element results was conducted. Failure load and load-deflection curves were used in the comparison.
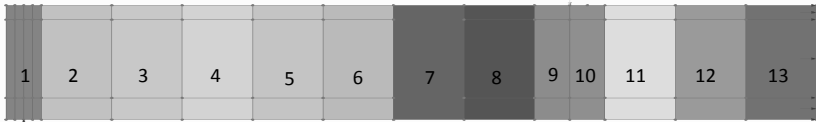
Fig 1. Zone division of the RC Beam for the FEA model
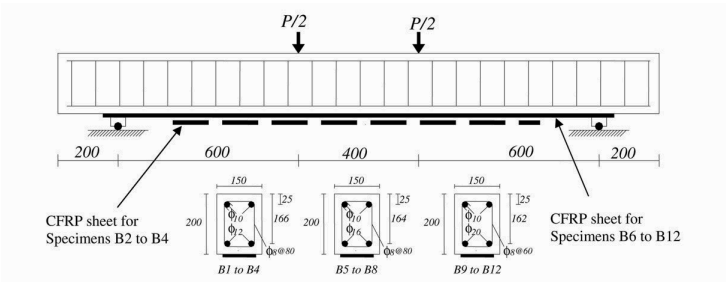
Fig 2. Esfahani RC beam details (Esfahani 2007)

Fig 3. Shows a comparison of load deflection graphs for the following cases:
(1) The experimental results; (2) the results for the non-linear FE model using a constant elastic modulus value for all zones; (3) FE results for elastic modulus obtained from the GA model updating process raw data; and modified results from the GA after visual inspection and rationalisation.
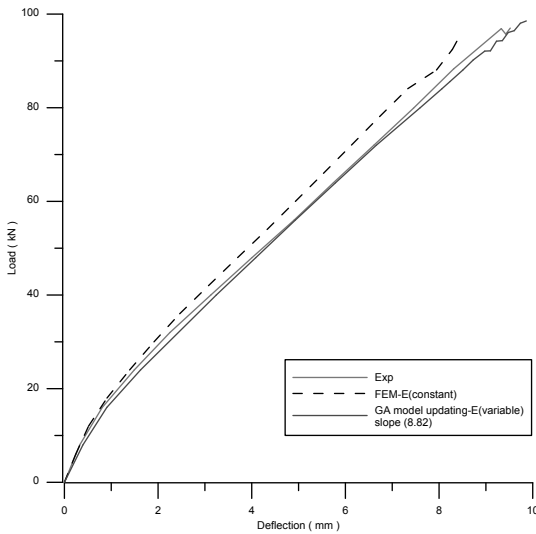
Fig 3 Load-Deflection curve for beam B9-20D-0L (20mm bar diameter)

A close inspection of the GA results revealed that the elastic modulus values varied along the length of the beam (from 22000 N/mm^2 to 12000 N/mm^2). At the zones close to the beam support, the elastic modulus values were increased to 22000 N/mm^2. This is in line with the findings of Rafiq and Sui (2010) that for the masonry panels in which the corrector factors around the panel boundaries were increased. One reason for this increase is that in FE modelling the boundaries are assigned as frictionless roller, while in reality this ideal support

type in not possible as there is always a degree of fixity at the supports. This fixity has been translated to a rise in the value of the elastic modulus at zones adjacent to the beam supports.

Similarly, values of the elastic modulus near the beam centre have reduced to about 12000 N/mm^2. The reason for this reduction could be that tensile (bending) crack appears early at these zones which leads to loss of stiffness at these zones. Hence it is realistic to decrease the elastic modulus at these zones.

To rationalise the elastic modulus generated by the GA, the beam was divided into two sections: one from the support to the point load; and the second from the point load location to the beam centre. Justification for this rationalisation process are: (1) the GA elastic modulus values mainly changed at these locations; (2) the section between the support and the point load are the regions where shear values vary; while at the section between the load and the centre of the beam there is no shear force, i.e. this section is pure bending regions. The elastic modulus for the first section was selected to be 22000 N/mm^2 and for the rest of the beam it was selected to be 12000 N/mm^2. The load deflection curve produced from this rationalised process is shown as GA-model updating-E in Figs. 3, 4 and 5. From these figures it can be seen that the gradients of the load deflection curves derived from modified model updated elastic modulus values are very similar to those obtained from the laboratory experimental results.
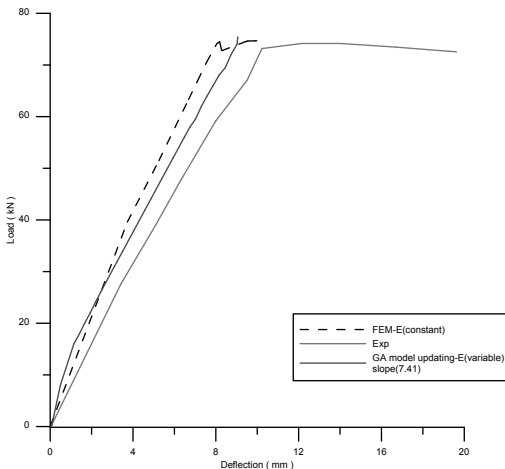
Fig 4 Load-Deflection curve for beam B5-16D-0L (16mm bar diameter)

It is worthwhile to note that in the GA optimisation process only the beam B9-20D-0L (20mm bar diameter) was used to find a set of elastic modulus values for the FEA non-linear analysis. The load deflection curve for this process is shown in Fig 3.

To test the generality of our findings the same elastic modulus values were used for two other beams [B5-16D-0L (16mm bar diameter) and B1-12D-0L (12mm bar diameter)] and the results of the analysis for these beams are presented in Figs 4 and 5.

This analysis demonstrates that the set of elastic modulus values found by the GA is not unique for the beam used in the GA optimisation process, but can be generally used for all beams with the same concrete properties, tested under 4 point load.
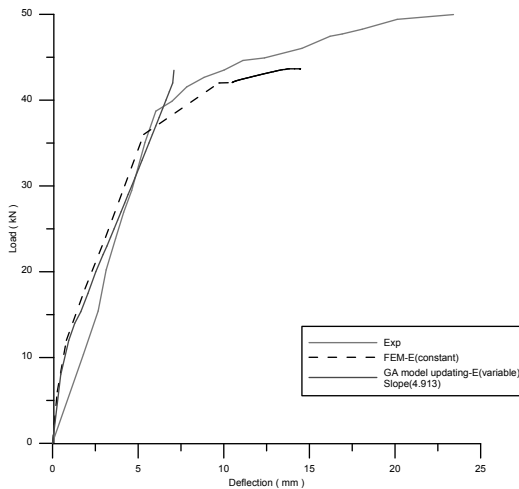
Fig 5 Load-Deflection curve for beam B1-12D-0L (12mm bar diameter)

Figs 4 and 5 show load deflection for beams B5-16D-0L B1-12D-0L. In these Figs the reinforcement bar diameter is 16mm and 12mm respectively.

Naturally a reduction in the bar diameter proportionally decreases the stiffness of the beam. Hence the load deflection curve gradient will be less than that of Fig 3. The gradient for the 20mm bar is 8.82 while the gradient for the 16mm and 12mm bars are 7.41 and 4.91 respectively. This gives us confidence in the model updating results of the GA.

From Figs 3 to 5 it is clear that the variation of the GA generated load deflection curves from the experimental curve is mainly due to variation at the region where there is no crack in the beam (i.e. a small region at the start of the curve).

The authors are in the process of testing the validity and generality of their findings, on beams of different dimensions, with different concrete properties and tested under the same 4 point method to investigate if their findings are generally applicable on beams with different properties.

CONCLUSIONS

- The GA successfully found a set of elastic modulus values for the concrete that resulted in a perfect match with the experimental results of the tested beam B9-20D-0L.
- When this model was used with two other beams (B5-16D-0L and B1-12D-0L), with bar diameters of 16mm and 12 mm respectively, the FEA load deflection curves slightly deviated from the experimental results. A close look at the results revealed that this deviation was due to the changes in the sizes of the steel reinforcement bars, which directly affects the stiffness of the beams. This was an added reassurance for the model used.
- A comparison of this model with that of the masonry panels confirmed that assuming standard boundary conditions for the beam can lead to an erroneous result. The GA successfully identified this and increased the rigidity of the model at zones near the supports by increasing the concrete elastic modulus values at these. Similarly the rigidity

of the beam was reduced near the centre of the beam, as the cracks appear first at these regions.

FURTHER WORK

- In the study presented in this paper, the model updating process was applied only to reinforced concrete beams with static loads.
- Further research by the authors is planned to look at the effects that CFRP has on the strength and load carrying capacity and what parameter should be considered in the optimisation process to get a best fit with the experimental results.
- This research will also be looking at the impact load and the FE models updated optimised parameters.
- Finally, we will be extending our investigation to damaged beams strengthened with CFRP and this would be a major contribution of this PhD programme.

REFERENCES

Aktan, A.E. Farhey, D.F. Brown, D.L. Dalal, V. Helmicki, A.J. Hunt, V.J. Shelley, S.J. (1996). "Condition assessment for bridge management", J. Infrastructure. Systems, ASCE 2 (3), 108-117.

Aktan, A.E. Farhey, D.F., Helmicki, A.J, Brown, Hunt, V. J, Lee, K. L., and Levi, A (1997), "Structural identification for condition assessment: experimental arts", J. Struct. Eng. ASCE 123 (12), 1674-1684.

Aktan, A.E., Catbas, N, Turer, A. Zhang, Z. (1998). "Structural identification: analytical aspects", J. Struct. Eng. ASCE 124 (7), 817-829.

Brownjohn, J.M.W., Pin-Qi Xia, Hong Hao, Yong Xia (2001) "Civil structure condition assessment by FE model updating: methodology and case studies", Finite Elements in Analysis and Design, 37 761-775.

Brownjohn, J.M.W., and Xia, P.Q., (1999). "Finite element model updating of a damaged structure", The Seventeenth International Modal Analysis Conference, SEM, Kissimmee, FL, pp. 457-462.

Brownjohn, J.M.W., and Xia, P.Q., (2000). "Dynamic assessment of a curved cable-stayed bridge by model updating", J. Struct. Eng. ASCE 126 (2), 252-260.

Brownjohn, J.M.W., Magalhaes, F., Caetano, E., and Cunha, A. (2010). "Ambient vibration re-testing and operational modal analysis of the Humber Bridge", Engineering Structures, 32(8), 2003-2018.

Esfahani, M.R., Kianoush, M.R. and Tajari, A.R. (2007). "Flexural behaviour of reinforced concrete beams strengthened by CFRP sheets", Engineering Structures, 29(10), 2428-2444.

Friswell, M, Mottershead, J. E. (1995). "Finite Element Model Updating in Structural Dynamics", Kluwer Academic Publishers, the Netherlands.

Goulet, J.-A., Smith, I. F. C., Texier, M. and Chouinard, L.(2012). The effects of simplifications on model predictions and consequences for model-based data interpretation. 20th Analysis & Computation Specialty Conference, ASCE, Chicago, Illinois, USA, March 29-31, 2012.

Goulet, J., Kripakaran, P., and Smith, I. (2010). "Multimodel Structural Performance Monitoring." J. Struct. Eng., 136(10), 1309–1318.

Raphael, B., and Smith, I. (1998). "Finding the right model for bridge diagnosis." Artificial Intelligence in Structural Engineering, LNAI 1454, Springer, 308-319.

Rafiq, M. Y., and Sui, C. (2010). "Interactive visualisation: A support for system identification", Advanced Engineering Informatics, 24(3), 355–366.

Sui, C. (2007) Improving the Prediction of the Behaviour of Masonry Wall Panels Using Model Updating and Artificial Intelligence Techniques. School of Engineering, Faculty of Technology. Plymouth, The University of Plymouth.

Sui, C., Rafiq, M. Y. & Zhou, G. (2008), "Preliminary Study of the Effects of Boundary Parameters on the Analytical Results of the Laterally Loaded Masonry Wall Panels", Proceeding of the Intelligent Computing in Engineering Conf. edt (Rafiq, Borthwick & de Wilde), Plymouth, July.

Zhou, G. (2002). " Application of stiff/strength corrector and cellular automata in predicting response of laterally loaded masonry panels" PhD Thesis. School of Engineering. Plymouth, University of Plymouth.

Zhou, G., Rafiq, M. Y., Easterbrook, D. J. & Bugmann, G. (2003) Application of cellular automata in modelling laterally loaded masonry panel boundary effects. Masonry International, 16 104-114.

Zhou, G. C., Rafiq, M. Y., Bugmann, G. & Easterbrook, D. J. (2006) Cellular Automata Model for Predicting the Failure Pattern of Laterally Loaded Masonry Wall Panels. Journal of Computing in Civil Engineering, Volume 20, 375-441.

Definition and Implementation of Temporal Operators for a 4D Query Language

S. Daum and A. Borrmann

Chair of Computational Modeling and Simulation, Technische Universität München, Arcisstraße 21, 80333 München; PH +49 89 289-23047; FAX +49 89 289 25051; email: {daum,borrmann}@bv.tum.de

ABSTRACT

Currently, there is no technology available to validate 4D Building Information Models using formal methods of temporal and spatial analysis. We propose to fill this technological gap by providing a query language which provides dedicated spatial and temporal operators. In the presented approach, a building is digitally represented by an instance of the neutral data format Industry Foundation Classes (IFC). With reference to the temporal and spatial data, the promoted IFC class structure is optimized for the particular query functionality. On the basis of this adapted object model we make use of the Language-Integrated Query technology (LINQ) provided by Microsoft's .NET framework. As LINQ was originally developed for defining static queries which are evaluated at compile time, we have implemented a dynamic version called Live LINQ in order to allow the user to flexibly formulate queries and to process these during the runtime of the BIM application. The developed interface provides a powerful and easy-to use mechanism for the spatial-temporal analysis and verification of 4D building information models.

BIM QUERIES

The concept of Building Information Modeling (BIM) is based on the use of intelligent, machine-readable representations of a building over its entire lifecycle. Using these representations, it is possible to provide extended computational functionality such as interactive query evaluation involving spatial and temporal predicates. This technology is well known from database management systems. However, BIM data models have particular characteristics that necessitate special implementations to realize the desired query support. In its standard form, a building model combines 3D geometry information with a complex object structure of components and their interrelations. By capturing construction schedule data, the BIM model can be extended to include temporal information concerning the construction and/or installation of individual components. This turns a conventional BIM model into a 4D model (Fischer, 2003).

Today, these 4D models have to be manually checked for correctness, which is a laborious and error-prone process. Based on our previous work concerning the realization of a spatial query language (Borrmann & Rank, 2009a,b), in this paper we discuss their extension through the inclusion of temporal operators. The resulting spatio-temporal query language facilitates the computational analysis and validation of 4D building models. An example of a spatio-temporal BIM query is "Select all non-supporting walls which touch floor 1 and ceiling 1 and which have been completed after ceiling 1". This request combines the examination of temporal, spatial and semantic predicates. Furthermore, the query is an example for the validation of a 4D model. The result represents all non-supporting walls in one particular storey. The system presented in this paper is able to process this query.

As input for the developed spatio-temporal query functionality, we make use of the vendor-neutral product model Industry Foundation Classes (IFC). However, the IFC object model has a complex and extensive data structure, particularly with respect to the geometric and temporal properties of building components, which makes the formulation of spatio-temporal queries extremely complicated. We therefore propose to transform this into a streamlined data structure. Combined with an extended version of the Language Integrated Query (LINQ) library, the system provides end users with an easy-to-use syntax for the formulation of queries involving spatial, temporal and semantic information.

RELATED WORK

One of the main functionalities of geographical information systems (GIS) is the processing of spatial queries. Though most of the available commercial GIS products restrict the spatial query functionality to 2D space, there has been ongoing, intensive research into 4D data storage and appropriate query technology (Noh, 2004). At the same time, the different modeling approaches in GIS and BIM make the direct application of GIS query technology for Building Information Modeling impossible. Entities in geographical information systems are set up as separate objects. They are often brought in a common context solely by their spatial location. BIM by contrast promotes a hierarchical product model with complex relations between the diverse components (buildingSMART, 2012). In addition, the representation of temporal data in GIS is not suitable for analyzing the construction and lifetime intervals of components in a BIM construction schedule.

Current research in BIM addresses methods for model filtering and the extraction of valid model views. General approaches using standardized database technology (SQL) have to be distinguished from domain-specific implementations. The complex IFC data structure and the associated object-relational mapping can lead to problems in performance and scaling when relational databases are employed (Wiet, 2012). Domain-specific developments range from native queries in the EXPRESS-X language to schema-based approaches such as GMSD (Weise, 2003) and PMQL, the Product Model Query Language of the EuroSTEP Model Server. Using these approaches, the creation of even a simple query can become very cumbersome (Borrmann & Rank, 2010). BIMQL is a promising research project realized as part of the bimserver.org platform (Wiet, 2012). This language operates

directly on the native IFC data format. The approach of using query shortcuts is also employed in our proposed system for geometry and temporal information.

DEFINITION OF THE TEMPORAL OPERATORS

The available data of the construction schedule represents the anticipated construction times of components. The user of the proposed BIM query system can additionally store information about the estimated lifetime of building parts. This makes it possible for the system to handle temporal queries concerning the lifetime interval of components.
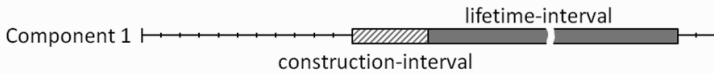
Figure 1: Construction interval and lifetime interval of component 1

To establish a common understanding of the meaning of the temporal operators, we institute a formal definition of their semantics based on the definitions provided by (Vilain, 1982). The temporal primitives employed here are intervals and points in time. The possible relations between these primitives are reflected by the 14 temporal predicates depicted in Table 1, which are implemented in our presented system. Using these predicates we are able to analyze the temporal relationships in a construction schedule.

Table 1: Temporal operators of the query language, based on (Vilain, 1982)

IFC DATA MODELING

The IFC standard includes about 600 classes, providing extensive data structures for the comprehensive and detailed modeling of buildings. IFC as a neutral data format has particular importance as it enables data exchange between software products from

different vendors, and serves as well as for realizing the hand-over of building data from the designers and contractors to the building operators.

The BIM query system we have developed uses ifcXML data conforming to the ifcXML2x3 standard as input (ifcXML2x3, 2012). A particular challenge is that the IFC make use of a very fine-grained object model involving a large set of inter-connected entities. In addition, the relationships between these entities are modeled as explicit objects. This makes the formulation of spatio-temporal queries using the original IFC schema extremely complicated. As a consequence, we propose adapting the model for the specific requirements of spatio-temporal analysis. This model resolves the objectified relationships and combines fine-grain data into more meaningful structures.

In the following, we describe our modifications of the IFC class structure to include temporal data. In the original model, scheduled construction intervals are represented by a branched structure composed of four different classes: The IfcRelAssignsToProduct class associates an IfcProduct, e.g. a wall, with a corresponding IfcTask. The task again refers to an IfcTaskTime, which finally includes a start and finish time (Figure 2).
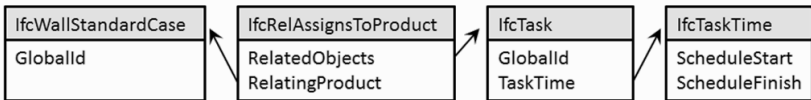
Figure 2: Relations between components and their construction times in the IFC

The data structure is simplified by adding a direct link between the IfcProduct object and the IfcTaskTime object. Thus, the recurring traversal of branches is not required for temporal queries. Furthermore, the relation arises at an IfcProduct which is a more natural starting point for use in a query than to examine a set of objectified relationships. The streamlined data structure is depicted in Figure 3.
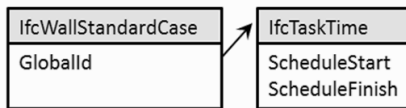
Figure 3: Simplified structure of components and their construction times

The geometrical information of IfcProduct objects is even more widely scattered across the IFC model. However, for the evaluation of the topological predicates explicit geometry is needed. Thus, the geometry information is transferred to explicit triangular meshes by use of external software. The resulting meshes are then linked to the appropriate IfcProduct objects.

SPATIAL AND TEMPORAL ALGORITHMS

The evaluation of the temporal algorithms is based on a comparison of the interval and points in time. For example, the start and end points are examined to compute

whether two intervals are overlapping. The speed of execution of these tests is fast enough to obviate the need for a supporting indexing structure.

However, the execution of topological operators is expensive and can slow down the system. To alleviate this a spatial index using an R* tree (Beckmann et al. 1990.) is constructed for IfcProduct objects. This eliminates many candidates using inexpensive pretests in topological queries. The topological algorithms are based on octrees. Here, the geometry of two IfcProduct objects is transferred on-the-fly to corresponding octrees. The cells of the tree, called octants, are classified by color attributes. The colors reflect the octant's spatial location concerning the hull of the original geometry. Possible color values are Black (Inside), Gray (Boundary) and White (Outside). By using a flooding approach, three colored octrees are produced. In the parallel traversal of two trees, color combinations are recognized. The 9 Intersection Model (9-IM) introduced by (Egenhofer, 1989) can now be used to deduce the accurate topological predicate. The collected color combinations of octants serve as input data. A detailed description of the implementation of topological operators is presented in (Daum & Borrmann, 2012).

FORMULATION OF FILTER EXPRESSIONS AND THEIR EVALUATION

In the proposed concept, a query expression is entered as source code by the end user or application programmer, respectively. The expression defines a predicate, which is used to select objects from a given set. The objects in the returned result set satisfy the predicate in the expression. By way of example, a set of walls is filtered by the expression wall.GlobalID.StartsWith("1pJQ") . Here, wall is a formal parameter which is replaced by all walls in the set one after another.

For the evaluation of query expressions at a set level, we use LINQ as it provides powerful query mechanisms for in-memory collections and object networks. LINQ is neatly integrated into the .NET framework and queries can be formulized by any .NET language. The queries are type safe and attributes and methods of involved objects can be used. For the definition of a query, an anonymous function, called a Lambda expression, like wall => wall.GlobalID.StartsWith("1pJQ") is used. The developed BIM system uses C# in combination with LINQ to examine in-memory collections. This means the filtering of building models can be executed directly in the fast main memory of the machine. When a 64-bit operating system is used, even very large building models can be queried without data having to be swapped to slower secondary storage. This is not possible with conventional database technology.

LIVE LINQ

When used as a standard component of .NET, LINQ requires the definition of filter predicates at compile time. While it is possible to pass parameters to a LINQ query at runtime, the system is restricted to pre-defined filters. To support flexible query statements that are entered by the end user at runtime, the expression must be converted to an executable filter object. This is achieved by on-the-fly compilation of user input at runtime. We call this functionality Live LINQ. If the entered C# code is accurate, it is compiled and saved as a .NET assembly, shown as a DLL in the file system. The query application can then load this assembly at runtime and use the

contained filter object in the pending filter execution. An unfiltered set of IfcProduct objects is supplied to the user as a data source. If a query is executed, the LINQ processor iterates over the IfcProduct collection. The returned set only contains entities, for which the dynamic filter evaluates to true.

As mentioned above, the IFC class structure can lead to complex filter statements. To make it easier to formulate filters, so-called query shortcuts are used and implemented with extension methods provided by the .NET framework. Extension methods make it possible to publish new methods on already existing types without having to recompile the original type. The approach of dynamic DLL creation and import is comparable to the query case: the difference is that the user enters a new extension method instead of a query. This method is compiled and loaded to the BIM application process. The extension method is automatically attached to the specific type. The user can simplify the query, as shown in Example 5, using a query shortcut.

In contrast to standard LINQ, the usage of Live LINQ allows the user to flexibly define queries. All object attributes can be used in a query and the system is not restricted to predefined filters. Additionally, as LINQ directly interact at an object level, methods published by objects are also accessible in queries, which make the temporal and topological operators available. Finally, filter generation at runtime means the compiler can be used as a validation tool for user-defined query statements. If inadequate code is submitted, the compiler returns meaningful error messages.

QUERY EXAMPLES

Example 1: Select walls with GlobalId 1pJQicIrH4UR_2BqvRP
```
IfcProducts.Where(p =>
      p is IfcWallStandardCase && p.GlobalId  == "1pJQicIrH4UR_2BqvRP")
```

Example 2: Select walls which overlap with wall 1 (index 25)
```
IfcProducts.Where(p => {
      var wall1 = IfcProducts [25];
      return p is IfcWallStandardCase  && p.Overlaps(wall1); })
```

Example 3: Select walls, if their construction-interval overlaps with that of wall1 and if they meet wall1
```
IfcProducts.Where(n => {
      var wall1 = IfcProducts [25];
      var constructionInterval1= wall1.ConstructionIntervall;
      return p is IfcWallStandardCase  &&
          p.TOverlaps(constructionInterval1) && p.Meets(wall1);})
```

Example 4: Select walls made of concrete
```
IfcProducts.Where(p =>
      p.hasAssociations[1].relatingMaterial.forLayerSet.
      materialLayers[1].material.name == "concrete")
```

Example 5: Select walls made of concrete (using the extension method)

```
IfcProducts.Where(p => p.GetMaterialName() == "concrete")
```
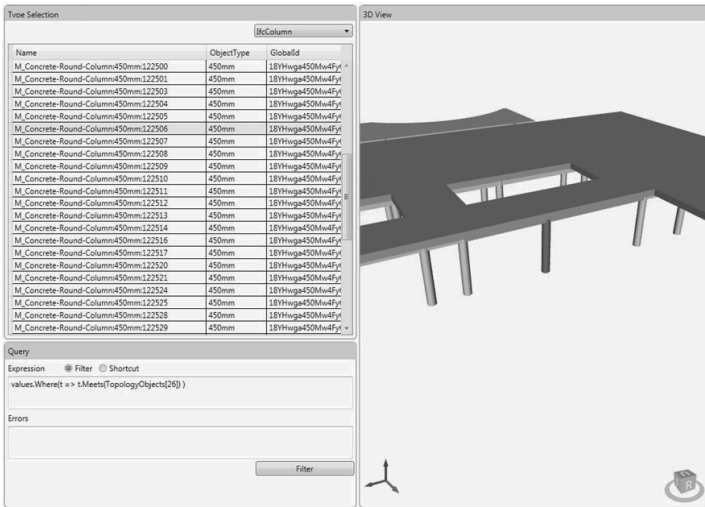

Figure 4: BIM system with spatio-temporal query functionality and LIVE LINQ

CONCLUSION

In this paper, we described the definition and implementation of a spatio-temporal query language for BIM. This language facilitates the computer-aided analysis of 4D models and the extraction of well-defined sub-models. As a basis for the query language we use the IFC standard, which provides a powerful modeling environment for the representation of buildings. Because the IFC class structure is only partially suitable for queries which directly involve spatial and temporal relationships, we propose a simplified data model adapted to meet the needs of representing and querying spatio-temporal properties and relationships. This significantly simplifies the formulation of query statements and reduces the computational costs of query evaluation. At the same time, the system is also capable of querying the original, complex IFC data structure. The queries are formulated as C# code and executed after an on-the-fly compilation in the LIVE LINQ environment. Object attributes and methods can therefore be included in query statements. In this way, spatio-temporal predicates are made available for direct query formulation. Using the described approach, we have been able to provide flexible BIM query functionality that provides spatial and temporal awareness.

REFERENCES

Beckmann, N., Kriegel, H.-P., Schneider, R., Seeger, B. (1990): The R*-Tree: An Efficient and Robust Access Method for Points and Rectangles. In: Proc. of the 1990 ACM SIGMOD International Conference on Management of Data.

Borrmann, A., and Rank, E. (2009a). "Topological analysis of 3D building models using a spatial query language" *Advanced Engineering Informatics* 23(4), 370-385.

Borrmann, A., and Rank, E. (2009b). "Specification and implementation of directional operators in a 3D spatial query language for building information models", *Advanced Engineering Informatics* 23 (1), 32-44

Borrmann, A., and Rank, E. (2010). "Query Support for BIMs using Semantic and Spatial Conditions" *Handbook of Research on Building Information Modeling and Construction Informatics: Concepts and Technologies,* IGI Global.

buildingSMART Ltd. (2012). "IFC data schemas" *Industry Foundation Classes* http://www.buildingsmart-tech.org/ifc/IFC2x3/TC1/html/index.htm> (10 January 2012).

Daum, S., and Borrmann, A. (2012). "Efficient and Robust Octree Generation for Implementing Topological Queries for Building Information Models" *Proceedings of the 19th EG-ICE International Workshop*, EG-ICE.

Egenhofer, M., Herring, J., (1989). "A mathematical framework for the definition of topological relationships." *Proc. of the 4th Int. Symp. on Spatial Data Handling.*

ifcXML2x3, (2012)."ifcXML reference", http://www.buildingsmart-tech.org/> (10 January 2012).

Fischer, M., Haymaker J., and Liston K. (2003) "Benefits of 3d and 4d Models for Facility Managers and AEC Service Providers", *4D CAD and Visualization in Construction,* Developments and Applications

Noh, S.-Y. (2004). "Literature Review on Temporal, Spatial, and Spatiotemporal Data Models." *Technical Report - Computer Science.* Iowa, USA, Iowa State University.

postGIS (2012). "Using PostGIS: Data Management and Queries", *postGIS reference* http://postgis.refractions.net/documentation/manual-2.0/using_postgis_dbmanagement.html#DE-9IM> (10 January 2012).

Vilain M. B. (1982): "A system for reasoning about time" AAAI-82 Proceedings

Weise, M., Katranuschkov, P., and Scherer, R. J. (2003). "Generalized model subset definition schema." *Proceedings of the 20th Conference on Information Technology in Construction.* CIB-W78.

Ontology-Based Building Information Modeling

Saeed Karshenas[1] and Mehrdad Niknam[1]

[1]Marquette University, Department of Civil, Construction and Environmental Engineering. Haggerty Hall, Milwaukee, WI 53233
Email:saeed.karshenas@marquette.edu, and mehrdad.niknam@marquette.edu

ABSTRACT

The World Wide Web has provided a vast distributed information space that is primarily designed for human consumption. The enmeshment of the Web and various business processes has made it imperative to define business information in a format that facilitates machine-to-machine exchange and automated processing. The construction industry is not an exception; construction business processes depend on a diverse set of information sources. For efficient processing of the construction business information, it is necessary to publish the information in a format that allows computer applications to easily discover, query and share the information.

Currently, the standard for representing, accessing, and sharing a building model is the Industry Foundation Classes (IFCs) data model. Several researchers have investigated the limitations of IFCs data model for interoperability among heterogeneous software applications. This paper presents the results of an on-going research on an ontology-based approach to building information modeling to facilitate information exchange among different knowledge domain applications. The approach developed in this study is based on a shared building ontology that models building domain element types and element relationships. The information systems to be integrated should be modeled using the shared ontology; each knowledge domain adds its own element properties to the shared building ontology. SWRL rules are used for mapping element properties from one domain to another. An example demonstrates how the developed approach can facilitate project information sharing between design and estimating domains.

Keywords: building modeling, ontology, interoperability, Revit, Cost estimating.

INTRODUCTION

Completing a construction project requires considerable efforts made with the input of a number of organizations such as architecture, engineering and construction companies, material and equipment suppliers, and labor organizations. Every organization involved in a project life cycle has an information system that needs to share information with other project participants. The World Wide Web has made it possible to make information systems universally accessible and these days almost all organizations are connected to the web. However, this has not solved the information sharing issues among various heterogeneous business applications.

For interoperability among a number of heterogeneous software applications, in a single enterprise or among networked enterprises, the applications must first agree on the

semantic of the information to be exchanged. In other words, the software systems must agree on a common information model. In computer and information sciences a common information model is called an ontology. According to Gruber (2009), an ontology specifies a set of representational primitives with which to model a domain of knowledge. The representational primitives are typically classes, attributes, and relationships. The specification takes the form of the definitions of representational vocabulary (classes, relations, and so forth), which provide meanings for the vocabulary and formal constraints on its coherent use.

This paper presents the results of an on-going research on an ontology-based approach to building information modeling to facilitate information exchange among different knowledge domain applications in the architecture, engineering, construction and facility management (AEC/FM) industry. To achieve semantic interoperability among heterogeneous applications, the meaning of the information that is exchanged must be understood across applications. This means semantic models of the information in the different application domains are needed. The approach developed in this study is based on a shared building ontology that models all element types and element relations in the building domain. The information systems to be integrated should be modeled using the shared building ontology. Semantic Web Rule language (SWRL, 2004) is used for mapping building element properties among knowledge domains. The application of this approach to information exchange between design and estimating domains is demonstrated.

RELATED WORK

The research efforts for capturing building domain knowledge and facilitating interoperability among architecture, engineering, and construction applications has led to standards such as STEP (ISO 10303-1, 1994) and Industry Foundation Classes (IFCs) (buildingSMART, 2010). There is also an XML representation of the IFCs schema called ifcXML that is designed for exchange of IFC data files as XML documents (ISO/TS 10303-28, 2003).

To introduce new means of distributed collaboration for the AEC/FM industry, Beetz, et al. (2006, 2009) developed OWL representation of the IFCs schema. They investigated two different approaches for deriving an OWL notation of IFCs: (1) OWL from XML schema of IFCs via XSLT and (2) OWL from EXPRESS schema of IFCs. The first approach was found unacceptable because of the excessive loss of model information while deriving the XML notation. Using the second approach they were able to capture most of the IFCs schema knowledge.

A number of researchers have discussed the shortcomings, and the reasons for low adoption of the IFCs data model (Eastman, et al., 2008, 2009; Olofsson, et al., 2008; Sacks, et al. 2010; Beetz, et al., 2009; Rezgui, et al., 20110). Venugopal, et al., (2012), describe IFC schema as rich but highly redundant, offering many ways to define objects, relations, and attributes which result in unreliable data exchanges.

To overcome the shortcomings of the IFCs model data for information interchange among AEC domain applications, a "use case" approach for specifying information exchanges has emerged (Eastman, et al., 2010). This approach involves the preparation of an Information Delivery Manual (IDM) by domain experts and development of Model View Definition (MVD) from IDM by IT experts. Venugopal, et

al. (2012) have improved this process by adding a semantic layer on top of the IFCs data model to eliminate ambiguities and clarify use cases.

To support interoperability among product development processes, the National Institute of Standards and Technology (NIST) has developed a Core Product Model (CPM) (http://www.mel.nist.gov/msidlibrary/doc/cpm_10_06.pdf). CPM is a generic model that is intended to be extensible and capture engineering context common to all product development activities. Dartigues etal. (2007) show the application of CPM for interoperability between a computer-aided design (CAD) system and a computer-aided process planning (CAPP) system. To achieve this objective, they modified CPM by adding some concepts that are common to both design and process planning domains. Next, the representation of different artifacts in the CAD and CAPP domains were analyzed to extract all the important concepts represented in each domain in order to correlate them with the common domain ontology entities. To achieve interoperability between the two domains they defined two algorithms: one to translate a file generated by the CAD software into instances of the common ontology and one to translate from the instances of the common ontology to a file that can be interpreted and processed by the CAPP software.

This paper presents a new approach to interoperability in AEC/FM industry. This approach is based on a shared building ontology that includes all element types used in various buildings and relations among the elements. Each AEC/FM domain builds its domain ontology on top of the shared ontology by adding appropriate properties to the element types defined in the shared ontology. There are a number of differences between NIST CPM and the shared building ontology. Among them are:

- The shared building ontology is a semantic model, CPM is a class hierarchy.
- The shared building ontology includes semantic definitions of all element types used in buildings; it is defined once and used as a base-level ontology for building other AEC/FM domain ontologies. CPM is a template for developing a common representation of objects in two domains.
- The shared building ontology is not directly used for mapping between two domains whereas CPM instances are part of the mapping mechanism between two domains.

The following sections discuss the shared building ontology and two separate domain ontologies (design and estimating domains) built on top of the shared ontology and a method developed for mapping an element's design properties to its estimating domain properties.

THE SHARED BUILDING ONTOLOGY

The methodology recommended by Uschold and King (1995) is used for ontology construction. This methodology includes four steps: purpose identification, ontology building, evaluation, and documentation. This section describes the main development activities and part of the shared ontology developed for buildings.

The purpose of the shared ontology is to create a base-level semantic model for representing buildings. The model includes only building element types and element relations. The shared building ontology must be created with input from all industry participants. This makes it possible to reuse the ontology to represent the knowledge of specific domains. Figure 1 shows a partial view of the developed shared ontology. The

full ontology includes all building element types; due to space limitation only two footing types are shown in Figure 1. The vocabulary and most classes used for building the ontology are borrowed from the IFCs schema; additional properties are added to show element relationships and classification. The classes and properties of the ontology are defined in OWL (W3C, 2004) RDF/XML syntax. The model in Figure 1 is a general, formal representation of a building model and provides a semantic foundation for modeling various domain views. A building instance includes a subset of elements in the shared building ontology. In the following sections the design and estimating views are discussed.
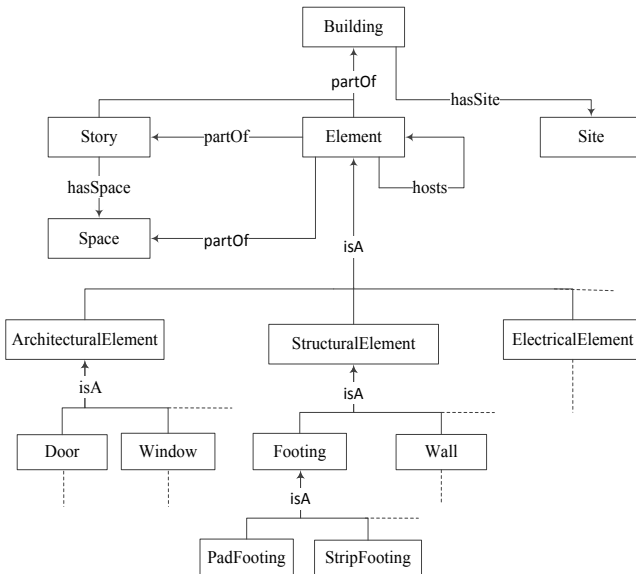
Figure 1 – Partial view of building domain ontology

THE DESIGN DOMAIN ONTOLOGY

Various AEC/FM domain ontologies are created by extending the shared building ontology shown in Figure 1. The design domain ontology contains specific knowledge related to the design of building elements. The extension of the pad footing in the design domain is shown in Figure 2. The design domain ontology adds element geometry and material properties to the shared ontology. A number of common element properties such as ID, name, and level are not shown in Figure 2.

Existing ontologies should be reused or extended when building a new ontology. The design domain ontology in Figure 2 uses two existing ontologies: (1) QUDT (http://www.qudt.org/) for representing units of measurement and (2) FC (http://www.freeclass.eu/freeclass_v1.html) to represent construction material. Quantities and units are the most important parts of building domain ontologies. Formal representation of measured data is important when data from several sources are

integrated. QUDT is an ontology of units of measurement. QUDT ontologies were developed for the NASA Exploration Initiatives Ontology Models (NExIOM) project, a Constellation Program initiative at the AMES Research Center (ARC).

FC ontology provides classes and properties for describing products and services from the building and construction industry. It is derived from the free classification standard freeClass. FC ontology is used in design domain ontology for representing building element material such as concrete, steel, and wood.

RDF uses only binary properties; therefore, to model multi-valued properties such as length, width, and material, auxiliary resources are used. The template used for naming the auxiliary resource for a property with length dimension is lengthSpec. So the width property of a pad footing has an auxiliary resource named lengthSpec that specifies the numerical value of the width property and its unit of measurement. Material property is specified by at least three attributes: material type, material quantity, and a unit of measurement. So, an auxiliary resource named materialSpec is defined with three properties.

Instances of design domain ontology are created from Autodesk Revit (http://usa.autodesk.com/revit/) model files. Revit model data are extracted using Revit API which provides direct access to the internal structure of a Revit model. The content
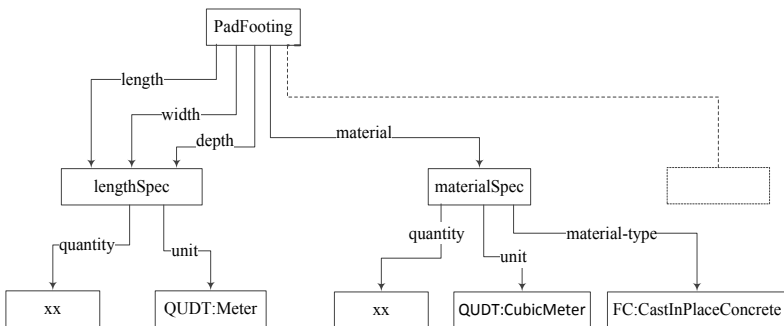
Figure 2- Partial view of the design domain ontology for a pad footing

of a CAD building model may be classified into five groups (Eastman, et al., 2009): functional type, geometry, attribute, relations between objects, and behavioral rules. Using Revit API, all model functional types, element geometry, element attributes and certain element relations can be accessed directly. Extracting some member relations require analyzing member geometries. In this research, extraction of behavioral rules is not investigated.

THE ESTIMATING DOMAIN ONTOLOGY

The estimating ontology contains specific knowledge related to the cost estimating domain. The estimating domain ontology is an extension of the shared building ontology. A partial view of the extension of the pad footing in estimating domain is shown in Figure 3. This ontology can produce responses to queries related to a

pad footing such as earthwork volume, formwork contact area, weight of rebar, or volume of concrete.
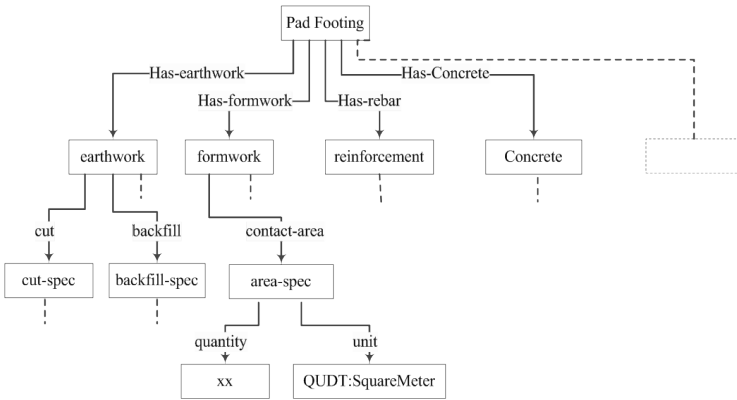
Figure 3- Partial view of the estimating domain ontology for a pad footing

SYSTEM ARCHITECTURE

The architecture of a semantic knowledge management system consisting of the design and estimating domains is shown in Figure 4. In this architecture, each domain has its own knowledge model that is created and maintained separately based on individual domain requirements. Each knowledge domain has its own knowledge editor for ontology management. In this study, Protégé 4.2 (http://protege.stanford.edu/) was used for domain ontology editing.

Each domain has repositories for its ontology and model instances that can be queried using SPARQL language (Prud'hommeaux, 2007). A user may access this type of application remotely and the knowledge and instance repositories may also be hosted on remote servers. In this study Sesame (http://www.openrdf.org/) database was used for saving ontology instances. The ontology middleware was implemented in Java language using Apache Jena and Pellet Reasoner.
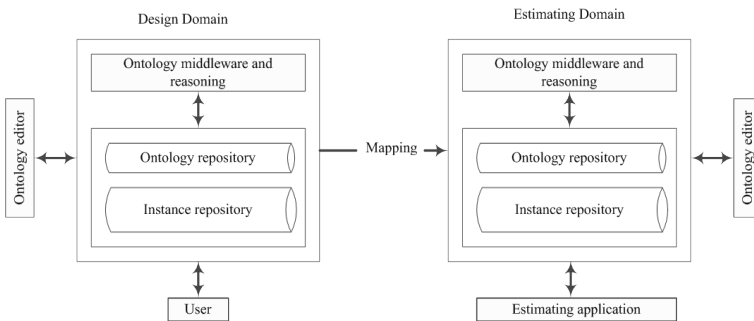
Figure 4- Semantic knowledge management architecture

In the above architecture, each knowledge domain has its own ontology that is built on top of the shared building ontology. The shared ontology contains the primitives of the building domain. It forms a sort of minimal skeleton of shared knowledge between the different domains. In the case of design and estimating domains, a series of SWRL (SWRL, 2004) rules may be used to determine estimating properties from design properties for an element instance. For example, for a pad footing with length = x, width = y and depth = z, formwork contact area may be assigned to the pad footing contact-area property using the following rule:

Formula for calculating contactArea = (2x + 2y)*z
SWRL rule: (Protégé 4.2 notations)
PadFooting(?p), length(?p, ?lenSpec1), width(?p, ?lenSpec2), depth(?p, ?lenSpec3),
     quantity(lenSpec1, ?x), quantity(lenSpec2, ?y), quantity(lenSpec3, ?z),
     add(?2x, ?x, ?x), add(?2y, ?y, ?y), add(?c, ?2x, ?2y), multiply(?r, ?c, ?z),
     hasFormwork(?p, ?form), contactArea(?form, ?area) -> quantity(?area, ?r)

Current BIM-based estimating applications require tedious mapping operations among building model elements and estimating software assemblies before element quantities can be transferred from the CAD model to the estimating software. The above mapping approach would eliminate the repetitive mechanical mapping work involved in the estimating process and greatly improves estimator's productivity.

SUMMARY AND CONCLUSIONS

This paper presented a novel approach for achieving semantic interoperability among heterogeneous information applications used in the building construction industry. The approach is based on a shared building ontology that models the element types and element relationships in buildings. Various domain applications that must interoperate should be built on top of the shared building ontology. This concept was demonstrated with an example for sharing pad footing information between the design and the estimating knowledge domains.

Development of a shared building ontology that is defined with the input from industry participants and is accepted within the AEC/FM industry would allow CAD software manufacturers to provide export options in their software for saving a building model in OWL format; similar to IFCs data model export option available in major CAD applications today. This would provide a major step towards solving the AEC/FM application integration and data sharing challenges. Moreover, basing industry data integration on ontologies would improve machine-readability of building data and their suitability to be used in a distributed environment.

REFERENCES
Beetz, Jakob, J. P. van Leeuwen, and B. de Vries, (2006). "An ontology web language notation of the industry foundation classes." *Proceedings of the 22nd CIB W78 Conference on Information Technology in Construction*. Vol. 2006. 2005.
Beetz, Jakob, Jos van Leeuwen, and Bauke de Vries, (2009). "IfcOWL: A case of transforming EXPRESS schemas into ontologies." *Artificial Intelligence for Engineering Design, Analysis and Manufacturing* 23.01 (2009): 89-101.

buildingSMART, (2010), Industry Foundation Class (IFC) data model, Web site,
    http://www.buildingsmart.org/.

Dartiges, C., Ghodous, P., Gruninger, M., Pallez, D., and Sriram, R. (2007). CAD/CAPP
    Integration using Feature Ontology. Cuncurrent Engineering: Research and
    Applications, Volume 15, Number 2.

Eastman, C. M., Jeong, Y. S., Sacks, R., Kaner I., (2009). "Exchange model and
    exchange object concepts for implementation of national BIM standards." *Journal
    of Computing in Civil Engineering* 24.1 (2009): 25-34.

Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2008). BIM Handbook: A guide to
    building information modeling for owners, managers, designers, engineers, and
    contractors. John Wiley & Sons Inc.

Gruber, T. (2009). "*Encyclopedia of Database Systems*", Ling Liu and M.
    Tamer Özsu (Eds.), Springer-Verlag.

ISO 10303-1, (1994). Industrial automation systems and integration – Product data
    representation and exchange – Part 1: Overview and fundamental principles.
    International Standards Organization 1994. TC 184/SC 4.
    http://www.iso.ch/cate/d20579.html.

ISO/TS 10303-28:2003 Industrial automation systems and integration – Product data
    representation and exchange – Part 28: Implementation methods: XML
    representations of EXPRESS schemas and data, International Organization for
    Standardization, Geneva, Switzerland, 2003.

Olofsson, Thomas, Ghang Lee, and Charles Eastman (2008). "Editorial-Case studies of
    BIM in use." *ITcon. v13* (2008): 244-245.

Prud'hommeaux, E., Seaborne, A.(2007). SPARQL query language for RDF.
    http://www.w3.org/TR/rdf-sparql-query/

Rezgui, Y., Boddy, S., Wetherill, M., & Cooper, G. (2011). Past, present and future of
    information and knowledge sharing in the construction industry: Towards
    semantic service-based e-construction?. *Computer-Aided Design*,*43*(5), 502-515.

Sacks, R., Kaner, I., Eastman, C. M., & Jeong, Y. S. (2010). The Rosewood
    experiment—Building information modeling and interoperability for architectural
    precast facades. *Automation in Construction*, *19*(4), 419-432.

SWRL: A Semantic Web Rule Language Combining OWL and RuleML, (2004). Web
    site, http://www.w3.org/Submission/SWRL/

Uschold, M., and King, M., (1995).
    http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf

Venugopal, M., Eastman, C. M., Sacks, R., & Teizer, J. (2012). Semantics of model
    views for information exchanges using the industry foundation class
    schema. *Advanced Engineering Informatics*.

W3C (2004). OWL Web Ontology Language. http://www.w3.org/TR/2004/REC-owl-
    features-20040210/.

A Semantic Web Service Approach to Construction Cost Estimating

Mehrdad Niknam[1] and Saeed Karshenas[1]

[1]Marquette University, Department of Construction, Civil and Environmental Engineering, Haggerty Hall, Milwaukee, WI 53233; email:mehrdad.niknam@marquette.edu and saeed.karshenas@marquette.edu

ABSTRACT

Cost estimating is a vital activity for contractors. Cost estimates require cost information collected from several sources. These sources include material and equipment suppliers. Current estimating software programs maintain a large database of resource cost data which must be updated before starting new estimates. The process of updating such databases is both time consuming and error prone. This paper discusses the architecture of a new approach to construction cost estimating using the semantic web technology.

The semantic web technology structures data according to formal ontologies intended to promote machine understanding. Semantic web services provide interfaces for publishing information. The proposed estimating approach requires that suppliers encapsulate their product information within appropriate interfaces and make this information available to contractors through web services. Estimating software applications can retrieve the latest cost data from suppliers' web services. The paper discusses how ontologies can be used to publish product information and develop ontology-based estimating applications.

INTRODUCTION

A building construction project is a collection of work items that produce building elements. To estimate the cost of a work item in a project, one needs to know the item quantity and the unit costs of the necessary resources. Current estimating applications keep a built-in database of resource (material, equipment, and labor) costs as shown in Figure 1. Item quantities are digitally extracted from a building model.

Resource costs continuously change; however, current estimating applications do not have the built-in capability to update their cost databases. Therefore, estimators must obtain the latest resource costs from suppliers and manually update estimating application databases, which is a time consuming process.
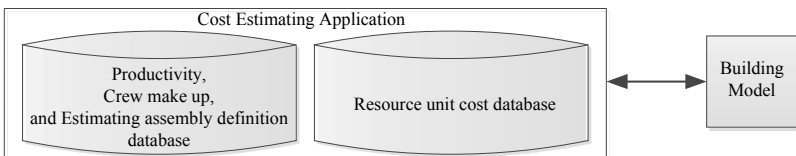
Figure 1. Current estimating application built-in databases

Providing direct access to the latest resource costs would eliminate the need for manual updating and has the potential to improve estimator efficiency. The semantic web services technology can be used by construction resource suppliers to semantically define and publish their resource offerings; thus, allowing estimating applications to access the latest resource costs when needed. Figure 2 illustrates this architecture. This approach allows estimating applications to internally store only information that does not change often such as crew make-ups, crew productivities, and estimating assembly definitions. Building model data may be accessed from a local or a remote service as shown in Figure 2.
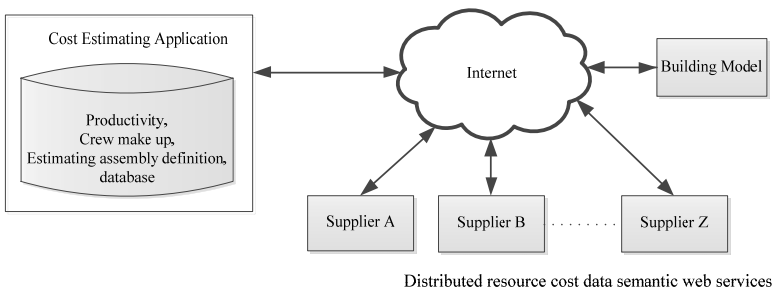
Figure 2. Proposed estimating architecture using semantic web services

A number of studies have focused on ontologies that describe construction processes (EI-Gohari and EI-Diraby 2010), construction material (http://semantic.eurobau.com), construction labor(Abanda et al. 2011) and building model (Staub-French, et al., 2003, Nepal, et al., 2012, another paper by the authors in this proceeding). However, there is no study on how these ontologies can be integrated into an estimating application. This paper describes a semantic web based estimating application that accesses material data published as semantic web services. It discusses how existing ontologies can be merged to develop an estimating domain ontology and tools that can be used for this purpose.

RELATED WORK

Enterprise applications are increasingly using service-oriented architecture to communicate with software services provided by different parties (Li et al. 2010). According to Martin et al. (2007), "The function of a web service is often to supply information, but can also involve exchange or sale of goods or obligations." Web services allow applications distributed over the Internet to communicate (exchange or produce new information) with each other.

Although service-oriented architecture facilitates data integration, it still requires human involvement because current web service technologies operate at the syntactic level (Roman et al. 2005). Web services support interoperability between application development platforms, but human involvement in discovering and understanding different functions of services is required. Web services do not provide the specification of what happens when a service is executed. Semantic descriptions can help automate web service discovery, composition, and execution (Fensel et al.

2011). Semantic data makes applications capable of understanding the meaning of the data.

The idea of the semantic web is to have a distributed web of data instead of a distributed web of pages (Allemang and Hendler 2011). Ontologies help define organization of the data distributed on the web by expressing a shared formal view between several parties (Noy and McGuinness 2001). Ontologies can be imported and used for knowledge representation. This gives computer applications awareness of the meaning of the data distributed on the Internet. Ontologies can be used to semantically describe web services. E-COGNOS (Lima et al. 2005) was a European project to develop ontology-centered web services, also known as e-processes, in construction domain (El-Diraby 2012). Domain ontology for construction concepts was developed as part of e-COGNOS (El-Diraby et al. 2005). Later, EI-Gohari and EI-Diraby (2010) developed ontologies for construction processes using e-COGNOS project experiences. Using ontologies to semantically describe web services and map them to business processes in construction projects allows a more automated integration process (Rezgui et al. 2011).

OWL-S is an ontology developed based on OWL to provide semantic web service description (Martin et al. 2007). It is a framework to provide computer-interpretable description of web services and the means by which they are accessed (Martin et al. 2004). Computer agents can use these descriptions to use web services. OWL-S provides semantic descriptions of web services by three main sub-ontologies of OWL-S: profile ontology, process model ontology, and grounding ontology. A service profile provides semantic descriptions of what the service does along with limitations, quality, and requirements of the service. Service-seeking agents use these descriptions to find the appropriate service that meets their needs. The process model describes how to use a service. These descriptions clarify how to request for a service and what happens when that service is requested. The process model can be used to compose and coordinate different web services. Service grounding specifies communication protocols, message formats, port numbers, and other details that describe how computer agents can access that service (Martin et al. 2004). This study uses OWL-S to semantically define supplier web services.

CONSTRUCTION PROJECT ONTOLOGY

A building project is a collection of processes that produce a building. Figure 3 shows the construction project process ontology used in this study. The project process ontology can be divided into four different categories based on functionality (EI-Gohari and EI-Diraby 2010). These categories are: management, knowledge integration, support, and core processes. Management processes ensure that design and construction follow the project objectives. Knowledge integration processes formally embed key concepts, knowledge, and experience into a project. Support processes (e.g. administration, communication, and information management) support other processes. Core processes create a project's primary products and include technical construction and design processes (El-Gohary and El-Diraby 2010).

In Figure 3, the technical construction process includes field activities that create building elements; these processes are the focus of this research. A technical construction process is composed of one or more sub-processes, also referred to as

work items that produce a building element. A work item is considered to be the smallest unit of work for estimating purposes (Schuette and Liska 1994). For example, to build a footing, the assembly of work items may include forming, reinforcing, and pouring.
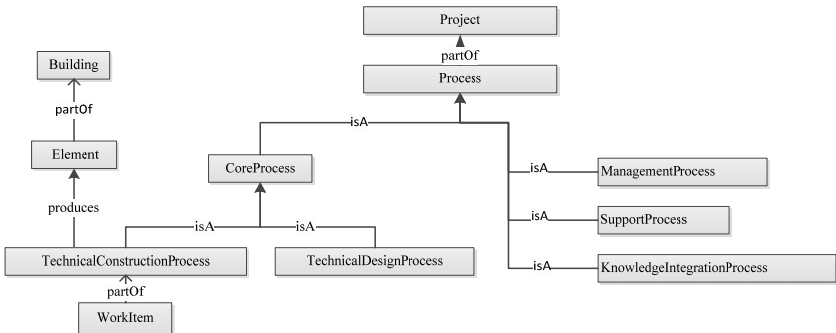
Figure 3. Construction process ontology

Figure 4 shows the ontology developed for work item in this study. In Figure 4, the organization concept represents the contractor responsible for performing the work item. Each work item has a schedule that shows the start and finish times, as well as the duration of the work item. A work item uses resources to produce a product. Multi-attribute properties such as quantity, has labor, and has equipment require special representations since RDF and OWL use only binary relations. For example to define a quantity, one must define a numerical value and a unit. Therefore, quantity specification class is used to provide definitions for quantity attributes. Similarly, labor specification would define labor type, unit cost, currency, and the number of workers needed in a crew.

The work item ontology uses existing ontologies defined by W3C (http://www.w3.org) for organization and time. Another existing ontology used in the work item ontology is FreeClassOWL ontology (FC) (http://semantic.eurobau.com) developed by the European Building and Construction Materials Database to describe construction materials and services. FC has over 60 million triples of real business data, 5676 classes, 174 properties, and 1423 instance data to describe construction materials. FC is W3C compliant. This paper does not discuss labor and equipment ontologies. Abanda et al. (2011) have developed domain ontology for labor cost knowledge and a rule-based application to facilitate labor cost calculations. The knowledge in their domain ontology is mainly focused on selecting a region, establishing different types of labor, their pay rates, and different languages they speak.
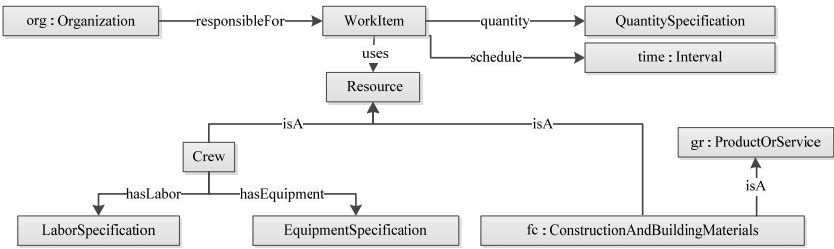
Figure 4. Work item ontology

The GoodRelations (GR) (http://www.heppnetz.de/projects/goodrelations) ontology, another existing ontology used in this research, allows businesses to provide descriptions about a product or service in RDF format. It has been widely used to annotate web commerce applications. In this research, GR is used to present an RDF description of construction material offerings. GR provides a vocabulary for describing products and services. In Figure 5, the ontology of a business entity offering a product or service is defined. As Figure 5 shows, a product or a service is described with qualitative and quantitative properties. A unit price specification describes the currency, currency value, and unit of measurement properties of an offer. GR allows business entities to provide more details about their offers such as available delivery methods, accepted payment methods, and advanced booking requirements.
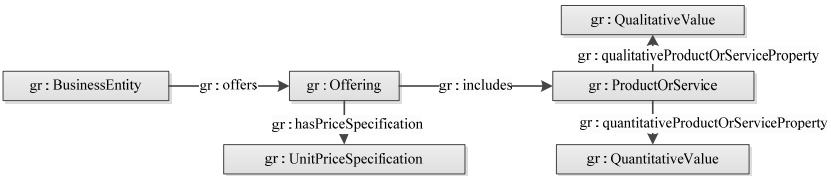
Figure 5. Part of Good Relations Ontology

FC and GR vocabularies allow material suppliers to clearly describe their construction product offerings and publish them as semantic services. Construction material suppliers can import and use these ontologies to define their products. This would allow estimating applications to understand the content of the data published by suppliers. An estimating application can retrieve the cost of a product by sending the product's specifications to the supplier's web service. This eliminates the current challenge of manually updating estimating application resource cost databases.

ESTIMATING APPLICATION IMPLEMENTATION

To develop an estimating application that implements the above approach, a number of software tools were used. These tools include Protégé 4.2(http://protege.stanford.edu) for building ontologies, Java programming language for developing application, Apache Jena and Pellet Reasoner (http://jena.apache.org)

for querying and reasoning OWL data, and Sesame repository (http://www.openrdf.org) for storing ontology instance data.

To create an estimate, the estimating application must access building model data for quantities, and supplier web services for resource costs. A product supplier web service was developed on a lab server; it offers a number of concrete products with different specifications. These products were published using FC and GR ontologies to provide supplier's web service semantics. A different paper in this proceeding discusses an ontology for building models.

The first step in creating the estimating application was to develop the estimating domain ontologies shown in Figures 3 and 4. A number of construction processes and their respective work items were defined. For example, "PadFootingProcess" is a process that produces the pad footings in a building and includes work items such as "Reinforcing", "Pouring", and "Forming". These processes were developed using Protégé 4.2. Protégé is an open source platform for creating and visualizing domain knowledge and supports OWL ontology language. It also allows importing and merging ontologies. Therefore, FC, GR, and building model ontologies were imported and merged with estimating domain ontology using Protégé.

Each work item in the estimating domain requires specific data from resource suppliers and the building model. For example, the "Pouring" work item requires concrete cost data from a concrete supplier and the volume of pour from the building model. Another example is the "Forming" work item that requires formwork cost data from a formwork supplier and formwork contact area from the building model.

The estimating application was developed using Java programming language and Apache Jena (2010). Jena is a framework for building semantic applications and supports SPARQL queries and rule-based inference engines such as Pellet Reasoner. To estimate the cost of an element in a building, the estimating application must be able to associate the element with the process that produces the element. A set of SWRL (http://www.w3.org/Submission/SWRL) rules are defined in the estimating ontology for this purpose. Once these rules are developed they can be imported and used in different projects. For example, the following rule maps all the pad footings in a building to the process that produces them:

Rule : PadFooting(?x) -> produces(PadFootingProcess, ?x)

This approach to estimating eliminates the need to manually map building model elements to estimating work items as is required when using current commercial estimating software.

Semantic web technology facilitates data integration in estimating applications. To estimate a work item, the estimating application communicates with the building model and a respective material supplier web service to obtain the necessary information. The data returned from these services are combined into a graph of knowledge about the work item being estimated. Figure 6 shows this process graphically. The element's RDF graph is saved in a Sesame data store. Sesame data store is an open source Java framework for storing and querying RDF data. Once element data is saved in the data store, the estimating application can query the data

store for the data needed to estimate the element cost. The estimated cost is also added to the element graph in the Sesame data store.
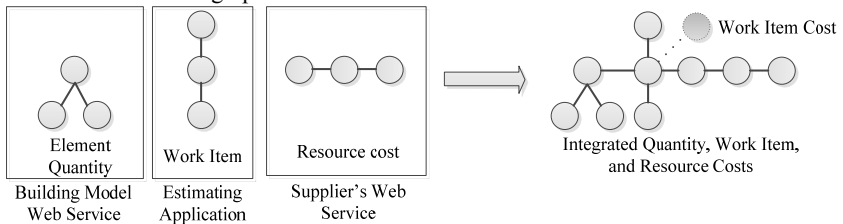
Figure 6. Integrating RDF data about a work item

SUMMARY AND CONCLUSION

This paper presented a novel estimating approach developed using ontologies, semantic web services, and a graph data store. In this approach, the estimating application directly communicates with resource supplier web services to retrieve up to date resource cost data. The application combines the retrieved resource cost data with quantities directly obtained from a building model to prepare an estimate. This eliminates the need for manually updating resource cost databases. This estimating methodology requires that suppliers publish their product data in a machine-interpretable format. To achieve this, suppliers must semantically define their products and publish their offerings as a semantic web service. One way to do this is to use FreeClassOWL (FC) and GoodRelations (GR) ontologies to semantically describe their products and services, and OWL-S to define their web services.

The presented approach to estimating is not an implausible idea. Semantic web service is a mature technology that is regularly used for web commerce. Currently, ontologies may not be used by construction material suppliers; however, several businesses have used ontologies to annotate their products and services. Examples include Best Buy and Overstock.com that have successfully implemented GoodRelations ontology to semantically describe their products. Google uses GoodRelations to enhance the information it provides in response to a search. So, the technology is readily available if material suppliers decide to use it.

The presented approach also requires digital access to building models. Currently, BIM-based applications can access building model information digitally. However, as discussed in the paper to eliminate the repetitive task of manually mapping model elements to estimating assemblies, an ontology-based representation of building models is required.

REFERENCES

Abanda, F., Tah, J., Pettang, C., and Manjia, M. (2011). "An ontology-driven building construction labour cost estimation in cameroon." .

Allemang, D., and Hendler, J. (2011). *Semantic Web for the working ontologist: effective modeling in RDFS and OWL.* Morgan Kaufmann, .

El-Diraby, T. E. (2012). "A Domain Ontology for Construction Knowledge (Dock 1.0)." *J.Constr.Eng.Manage.,* .

El-Diraby, T., Lima, C., and Feis, B. (2005). "Domain taxonomy for construction concepts: toward a formal ontology for construction knowledge." *J.Comput.Civ.Eng.,* 19(4), 394-406.

El-Gohary, N. M., and El-Diraby, T. E. (2010). "Domain ontology for processes in infrastructure and construction." *J.Constr.Eng.Manage.,* 136(7), 730-744.

Fensel, D., Facca, F. M., Simperl, E., and Toma, I. (2011). "Web Service Modeling Ontology." *Semantic Web Services,* 107-129.

Li, G., Muthusamy, V., and Jacobsen, H. A. (2010). "A distributed service-oriented architecture for business process execution." *ACM Transactions on the Web (TWEB),* 4(1), 2.

Lima, C., El-Diraby, T., and Stephens, J. (2005). "Ontology-based optimization of knowledge management in e-construction." *Journal of IT in Construction,* 10 305-327.

Martin, D., Burstein, M., Hobbs, J., Lassila, O., McDermott, D., McIlraith, S., Narayanan, S., Paolucci, M., Parsia, B., and Payne, T. (2004). "OWL-S: Semantic markup for web services." *W3C Member Submission,* 22 2007-2004.

Martin, D., Burstein, M., Mcdermott, D., Mcilraith, S., Paolucci, M., Sycara, K., Mcguinness, D. L., Sirin, E., and Srinivasan, N. (2007). "Bringing semantics to web services with owl-s." *World Wide Web,* 10(3), 243-277.

Nepal, M. P., Staub-French, S., Pottinger, R., & Zhang, J. (2012). Ontology-Based Feature Modeling for Construction Information Extraction from a Building Information Model. *Journal of Computing in Civil Engineering*.

Noy, N. F., and McGuinness, D. L. (2001). "Ontology development 101: A guide to creating your first ontology." .

Rezgui, Y., Boddy, S., Wetherill, M., and Cooper, G. (2011). "Past, present and future of information and knowledge sharing in the construction industry: Towards semantic service-based e-construction?" *Comput.-Aided Des.,* 43(5), 502-515.

Roman, D., Keller, U., Lausen, H., de Bruijn, J., Lara, R., Stollberg, M., Polleres, A., Feier, C., Bussler, C., and Fensel, D. (2005). "Web service modeling ontology." *Applied Ontology,* 1(1), 77-106.

Schuette, S. D., and Liska, R. W. (1994). *Building construction estimating.* McGraw-Hill, .

Staub–french, S., Fischer, M., Kunz, J., Ishii, K., & Paulson, B. (2003). A feature ontology to support construction cost estimating. *AI EDAM: Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *17*(02), 133-154.

Conceptualizing Methodology for Building an Ontology for Construction Claim Knowledge

Jia Niu[1] and Raja R.A. Issa[2]

[1]Rinker School of Building Construction, University of Florida, PO Box 115703, Gainesville, FL, USA32611-5703;PH(352)328-7483; email: jniu@ufl.edu
[2]Rinker School of Building Construction, University of Florida, PO Box 115703, Gainesville, FL, USA32611-5703; PH(352)273-1152; email: raymond-issa@ufl.edu

ABSTRACT

Sharing comprehensive and professional knowledge about construction contractual claims through ontology could be a good way to improve contractual management. However, when it comes to the implementation of the ontology development, regarding the characteristics of claim knowledge, the practical and feasible conceptualization methodology is not yet well developed. This paper explores a synthesized methodology to fulfill the conceptualization work for the domain knowledge of construction claims. Based on a literature review on knowledge sharing in the construction claims area and ontology building methodologies, as well as the target ontology's definition and scope, the conceptualization activity in the ontology developing methodology of METHONTOLOGY, was selected as a raw conceptualizing methodology to use in the validation case study. The case study was designed to conceptualize the claim knowledge contained in the AIA A201 General Conditions of Contract Document. The results of the case study indicated that METHONTOLOGY provides a valid framework and work flow for this domain.

INTRODUCTION

In today's construction industry, claims and legal issues have become more and more unavoidable due to the increasing complexity and uncertainty involved in projects (Hackett and Dancaster 2000). However, solid and reliable legal advice is not available and not convenient to obtain due to its expensiveness and the remoteness of job site locations. Therefore, sharing the domain knowledge about claims and legal issues among the project management team members would necessarily be beneficial in dealing with contractual claims and disputes. It has been widely accepted that ontology-based development can be an effective methodology for sharing domain knowledge. Many ontology-based studies have been conducted

in a variety of research areas, and a research framework supported by sharing and using the domain knowledge about construction contractual claims has been proposed (Niu and Issa 2012). But, when it comes to the implementation of such an ontology, the theoretical support and practical guide pertinent to this specific domain are both fairly limited. This paper proposes a practical and feasible methodology to fulfill the conceptualization work for building the target ontology, and validating its robustness through a case study.

LITERATURE REVIEW
Sharing Knowledge in the Legal Area

The significant applicability of knowledge sharing in the legal area was addressed nearly three decades ago. In 1980s, scholars noted that the law is a near-perfect application area for knowledge representation (Cross and deBessonet 1985). Specifically, in the construction contractual claim area, with the proliferation of expert systems in AI (Artificial Intelligence) field in 1980s, a number of studies on expert systems were initiated to undertake the legal analysis for certain scenarios (Arditi and Patel 1989; Cooper 1994; Diekmann and Kim 1992; Kraiem 1988). However, most of those expert systems were developed on the one same principle for knowledge representation, using production rules scheme.

Ontology Building Methodologies

Another meaningful result of AI's development in the 1980s is that, the community began to realize the importance of capturing knowledge for powerful AI system, which leads to the initial usage of the term ontology to refer to the modeled body of knowledge. Gruber (1993) first coined the term of "ontology". After that, a large amount of research regarding ontology from theory to methodology was initiated.

As far as the methodology for building ontology is concerned the first guidelines for developing ontologies, ENTERPRISE Ontology and TOVE (Toronto Virtual Enterprise), were proposed by Uschold (1995) and Grüninger and Fox (1995) respectively and refined later (Uschold 1995; Uschold and Grüninger 1996). However, these guidelines were only applicable to the case studies involved and did not set of standard for all situations.

Actually, depending on the characteristics of different domain areas, the process of building ontologies can vary. Thus, more theories about the methodology of ontology building were developed under different background. The influential ones include: KACTUS methodology (Bob et al. 1994; Guus et al. 1995) for reuse of knowledge about technical system during their life-cycle in the manufacturing and engineering domains; IDEF5 (Integrated Definition for Ontology Description Capture Method) method to develop and maintain usable, accurate, domain ontologies in software engineering (Benjamin et al. 1994); METHONTOLOGY for

building ontology for chemistry knowledge (Lopez et al. 1997; Lopez et al. 1999); guideline and methodology from natural language based SENSUS Ontology for machine translation (Swartout et al. 1997); On-To-Knowledge methodology for knowledge management systems (Staab et al. 2001); Noy and McGuinness (2002) methodology consisting of seven steps.

METHODOLOGY
Definition and Scope of the Target Ontology

As mentioned earlier, in order to share construction contractual claim knowledge through an ontology-based system, the first task is to define the desired body of knowledge. Furthermore, an analysis of exactly what kind of knowledge is actually needed to achieve the goal of conducting a claim analysis and then providing legal advice predicting both the risk and opportunity for a claim arising. For instance, assuming the user is a general contractor, then the risk represents the odds that the owner would make a claim on the general contractor; whereas the opportunity is the possibility that the general contractor would make a claim for additional time and/or cost on the owner. Additionally, the risk and opportunity for making a claim comes from the contractual rights and obligations agreed to by both parties.

Base on this assumption, therefore, knowledge about the causal mapping between the reasons and results is required. Here the reasons represent the causes that could incur any claim events. For example, "Price Escalation", "Design change", "Failure of Owner's work" and any scenario that could lead to a claim event. For this part, collecting and classifying construction claim cases would provide a valid empirical clue (Chehayeb and Al-Hussein 2005). Further, the desired knowledge should include the comprehensive and exhaustive categories of claim-incurring causes.

In addition, another body of knowledge that should be taken into consideration is the decision process used by a lawyer in analyzing claims. Generally, this process mainly consists of consideration of applicability of contract conditions and clauses; identification of certain patterns of fact as claim event causes; and evaluation of the robustness of the claim arguments. To make this decision process work, certain sub-knowledge is necessarily needed, which includes standards for judging the validity of the facts and technical information as a claim cause. The handling method for uncertainties, for example, how to deal with the legal terms such as reasonable and materially that often appear in contract documents, is also a critical component.

Conceptualization Methodology Adopted

Based both on the comparison framework which considers factors such as ontology construction strategy, software tool support, ontology development processes, and application experience, and the IEEE standard for software

development (Fernández López 1999; Fernández López et al. 2002), METHONTOLOGY has a high degree of maturity. To be specific, it has a more clear work flow and detailed plan, which makes it more practical and feasible than other methodologies, and fits the need of this study focusing on hands-on operation. Corcho et al. (2005) used METHONTOLOGY to develop a legal ontology domain for general Spanish law. However, the ontology because of its generality is not valid for the construction claim domain.

The conceptualization activity in METHONTOLOGY organizes and converts an informally perceived view of a domain into a semi-formal specification using a set of intermediate representations (IRs) based on tabular and graph notations that can be understood by domain experts and ontology developers. For the conceptualization activity, METHONTOLOGY includes the set of tasks for structuring knowledge, as shown in Figure 1 (Gomez-Perez et al. 2004). The case study follows this process to conduct the conceptualizing work and validate the derived process.
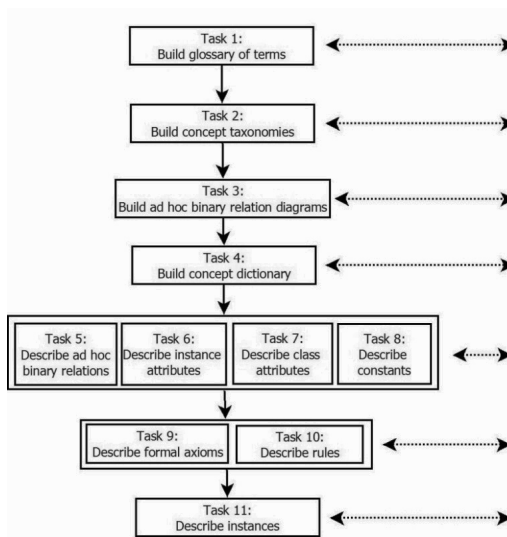
Figure 1. Tasks of the conceptualization activity according to METHONTOLOGY (Adapted from Gomez-Perez et al. 2004)**.**

CASE STUDY: CONCEPTUALIZING AIA A201

As a widely accepted construction contract document by the academia and industry, AIA documents serve as a comprehensive and professional representation of the contractual relationships between different construction project parties. One of the most common and typical contractual relationships is the one between the Owner

and the Contractor, the General Conditions of Contract (AIA A201 2007) document is selected as the knowledge source for this case study. The goal of this case study is to use the methodology of the conceptualization activity in METHONTOLOGY to acquire and conceptualize the claim knowledge from the AIA A201 document. Due to the length limit of this article, only Task 1-3 here, which are the most crucial ones in this conceptualization methodology will be explained.

For the execution plan, the starting point is to building the glossary of terms (Task 1). Fortunately, the well-organized "Index" part of the AIA A201 document makes a good reference to facilitate this initial step, since this part is a glossary for the document itself for the reader to find related clauses which contain a certain term. About 300 terms listed in the "Index" part are included in the glossary of terms. An excerpt of the glossary of terms is shown in Table 1.

Table 1. Excerpt of Glossary of Terms

Name	Synonyms	Acronyms	Type	Description
Acceptance of Nonconforming Work	Nonconforming Work, Acceptance of	--	Concept	Owner accept Work that is not in accordance with the requirements of the Contract Documents instead of requiring its removal and correction
Acceptance of Work	--	--	Concept	Owner accept Work
Access to Work	--	--	Concept	Access to the Work in preparation and Progress wherever located
Accident Prevention	--	--	Concept	Duty of prevent accidents
Acts and Ommisions	Action and inaction	--	Concept	The legal behavior of a certain subject
Addenda	--	--	Concept	A part of Contract Document, which is for additional material added at the end of it
Addtition Cost, Claims for	Claims for Additional Cost	--	Concept	A kind of claim asking for cost
Additional Inspections and Testing	--	--	Concept	Inspections and testings occured by the rejected Work
Additional Time, Claims for	Claims for Additional Time	--	Concept	A kind of claim asking for time
Administration of the Contract	--	--	Concept	A kind of duty
...	...	...	...	...

Once the glossary of terms is built, then the following steps are to build concept taxonomies to classify concepts (Task 2), and then build the ad hoc binary relation diagrams (Task 3) to identify ad hoc relationships between one or more taxonomies. For example, as shown in Figure 2, the taxonomy for the concept Contract Document represents that the Contract Document consists of Agreement, Conditions, Drawings, Modification, Addenda, Specifications and other document, and those six parts are disjoint from each other and cover Contract Document fully, which is a *Partition* of the concept Contract Document. Additionally, the arced arrow lines in Figure 2 represent ad hoc relationships, e.g. Contract Document does not include the information for bid, and vice versa. Additionally, by using the taxonomies and ad hoc binary relationships, more conceptual models representing the semantics contained in the contract articles and clauses can be extracted.

Problems Encountered and Analysis

With the progress of the case study, the terms in the glossary of the "Index" part of the AIA A201 document were found not to be the exact glossary that was
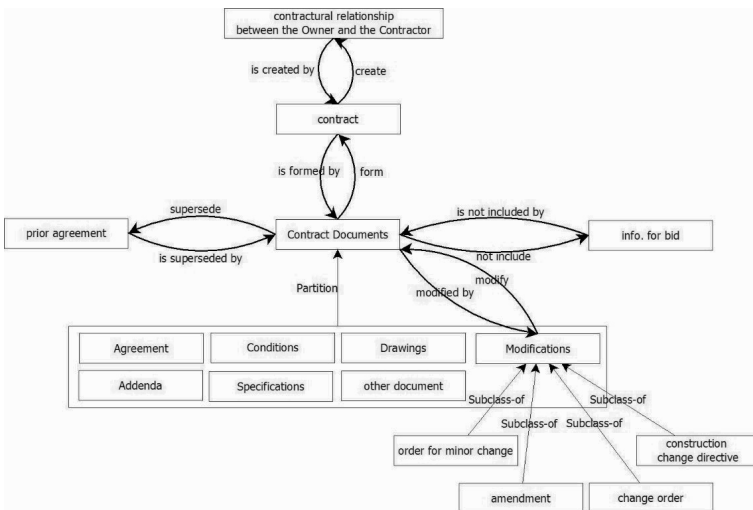
**Figure 2. Excerpt of Taxonomy and Ad hoc Relationships
of Construction Claim Ontology**

needed to build claim knowledge, most of them needed to be modified for further use. For example, the interchangeable terms were combined as one (e.g. Completion, Substantial and Substantial Completion are the same); those terms serving only as a reading guide but without any actual semantics were discarded (e.g. Basic Definitions, Capitalization); additionally, many terms actually contained a cluster of concepts and relations, like a "block" rather than one single term, such as Commencement of the Work, Conditions Relating to, Defective or Nonconforming Work, Acceptance, Rejection and Correction of, Shop Drawings, Product Data and Samples and so on. For these terms, it is necessary to go deep into the "block" in order to conceptualize them (e.g. Conditions relating to … terms) or to generalize those independent concepts into one as a superclass if they have similar characteristics or behaviors for claim purposes (e.g. Shop Drawings, Product Data and Samples).

Another problem encountered was how to classify concepts in the taxonomy building phase. To be specific, it is usually confusing to classify some concept as a "class" or "instance", and "concept" or "relation". For example, the term of Advertisement or Invitation to Bid could either be a kind of information for bid as a subclass, or an instance of it. On the other hand, the term Acceptance of work could be deemed as a relation between the Owner and the Contractor, but it also makes sense to treat it as a concept which represents the Owner's intent to give up the right of requiring removal and correction. Noy and McGuinness (2002) in their methodology pointed out these kind of situations and provided corresponding rules and guidelines for handling them.

CONCLUSION AND FUTURE WORK

　　Through the case study, the suitable conceptualization methodology for building domain ontology for construction claim knowledge was explored. METHONTOLOGY provides a general framework and work flow for building domain ontology, and it can also be adopted in developing the domain knowledge about construction claim. In addition, Noy and McGuinness' methodology (2002) about defining classes and a class hierarchy (taxonomy) provide a helpful guideline to deal with some specific term classifying issues in developing the taxonomies. It would be a better strategy to synthesize these two methodologies for further ontology building efforts in the construction claim area. Moreover, further conceptualization work, especially, in building the taxonomy is in great need of solid expertise and keen insight from domain experts to better define and organize concepts. The validation method developed for the ontology should focus on the interaction with domain experts as well as be in accordance with precedent law cases. Future efforts should focus on developing a practical and feasible validation method.

REFERENCES

Arditi, D., and Patel, B. K. (1989). "Expert system for claim management in construction projects*." International Journal of Project Management*, 7(3), 141-146.

Benjamin, P. C., Menzel, C. P., Mayer, R. J., Fillion, F., Futrell, M. T., deWitte, P. S., and Lingineni, M. (1994). "IDEF5 Method Report." Knowledge Based Systems, Inc.

Bob, W., Guus, S., Wouter, J., Anjo, A., and Harmelen, F. v. (1994). "Framework and Formalism for Expressing Ontolgies." University of Amsterdam.

Chehayeb, A., and Al-Hussein, M. (2005) "Canadian Construction Claim Tracker: An Integrated Repository System for Canadian Case-Law Construction Claims." *Construction Research Congress 2005*, ASCE, 116.

Cooper, T. E. (1994). "A knowledge-based construction claims advisor for the AIA A201 General Conditions document." Thesis (PhD), Auburn University, Auburn, AL, United States.

Corcho, O., Fernández-López, M., Gómez-Pérez, A., and López-Cima, A. (2005). "Building Legal Ontologies with METHONTOLOGY and WebODE." *Law and the Semantic Web*, V. R. Benjamins, P. Casanovas, J. Breuker, and A. Gangemi, eds., Springer Berlin Heidelberg, 142-157.

Cross, G. R., and deBessonet, C. G. (1985). "Representation of legal knowledge for conceptual retrieval." *Information Processing & Management*, 21(1), 35-44.

Diekmann, J. E., and Kim, M. P. (1992). "Supercharge: expert system for analysis of changes claims." *Journal of Construction Engineering and Management*, 118(2), 399-410.

Fernández López, M. (1999). "Overview of Methodologies For Building

Ontologies." *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Methods KRR5 Stockholm*, Sweden, August 2, 1999.

Fernández López, M., Gómez Pérez, A., and Oacute (2002). "Overview and analysis of methodologies for building ontologies." *The Knowledge Engineering Review*, 17(02), 129-156.

Gomez-Perez, A., Corcho, O., and Fernandez-Lopez, M. (2004*). Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*. First Edition (Advanced Information and Knowledge Processing), Springer.

Gruber, T. R. (1993). "A Translation Approach to Portable Ontology Specification." *Knowledge Acquisition*, 5, 199-220.

Grüninger, M., and Fox, M. (1995) "Methodology for the Design and Evaluation of Ontologies." *Proc., IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing*, April 13, 1995.

Guus, S., Bob, W., and Wouter, J. (1995). "The KACTUS View on the 'O' Word." *International Joint Conference on Artificial Intelligence*.

Hackett, J., and Dancaster, C. (2000). *Construction claims: Current practice and case management*, LLP, London.

Kraiem, Z. M. (1988). "DISCON: An expert system for construction contract disputes." Thesis (PhD), Colorado University, Boulder, United States.

Lopez, M., Perez, A., and Juristo, N. (1997). "METHONTOLOGY: from Ontological Art towards Ontological Engineering." *Proc., Proceedings of the AAAI97 Spring Symposium*, 33-40.

Lopez, M. F., Gomez-Perez, A., Sierra, J. P., and Sierra, A. P. (1999). "Building a chemical ontology using Methontology and the Ontology Design Environment." *Intelligent Systems and their Applications*, IEEE, 14(1), 37-46.

Niu, J., and Issa, R. R. (2012). "Framework for Production of Ontology-based Construction Claim Documents." *ASCE International Conference on Computing in Civil Engineering*, R. R. Issa, and I. Flood, eds., ASCE, Clearwater Beach, Florida, 9-16.

Noy, N. F., and McGuinness, D. L. (2002). "Ontology Development 101: A Guide to Creating Your First Ontology." *Knowledge System Laboratory*, Stanford, CA.

Staab, S., Studer, R., Schnurr, H. P., and Sure, Y. (2001). "Knowledge processes and ontologies." *Intelligent Systems*, IEEE, 16(1), 26-34.

Swartout, B., Ramesh, P., Knight, K., and Russ, T. (1997). "Toward Distributed Use of Large-Scale Ontologies." *AAAI Symposium on Ontological Engineering*.

Uschold, M. (1995). "Towards a Methodology for Building Ontologies." *Workshop on Basic Ontological Issues in Knowledge Sharing IJCAI*.

Uschold, M., and Grüninger, M. (1996). "Ontologies: principles, methods, and applications." *Knowledge Engineering Review*, 11(2), 93-155.

A Tool for Automatically Tracking Object Changes in BIM to Assist Construction Managers in Coordinating and Managing Trades

Brittany Giel[1], Raja R.A. Issa[2], Rui Liu[3] and Le Zhang[4]

[1]Rinker School of Building Construction, University of Florida, PO Box 115703, Gainesville, FL, USA 32611-5703; PH (352) 273-1178; email: b1357g@ufl.edu
[2]Rinker School of Building Construction, University of Florida, PO Box 115703, Gainesville, FL, USA 32611-5703; PH (352) 273-1152; email: raymond-issa@ufl.edu
[3]Construction Science and Management Program, College of Architecture, 501 W. César E. Chávez Blvd, San Antonio, TX -78207; PH (210) 458-3054; email: rui.liu@utsa.edu
[4]M-SIX Technology, Inc. Portland, OR 97209; PH (352) 949-9419; email: le.zhang@m-six.com

ABSTRACT

As Building Information Modeling (BIM) becomes more popular in the construction industry, one of its most useful benefits to construction managers is the assistance it provides them in coordinating different building trades during preconstruction. Tools such as Autodesk's Navisworks for conflict detection have been instrumental to improving the speed, efficiency and accuracy of the coordination process. However, even with these tools, construction managers are faced with a daunting task of managing multiple versions of different trades' models and communicating changes downstream to those responsible for modeling. This paper proposes the development of a tool to evaluate the productivity of model updates made by different trades for different projects. The tool automatically generates a report for 3D object changes between different model versions. Through case study analysis, the tool was validated for the Mechanical Duct trade of a sample building project. Preliminary results indicate that it may aid construction managers in assessing the modeling productivity of different trades and making predictions regarding the estimated duration of the coordination phase of a project.

INTRODUCTION

Construction Trade Coordination refers to the physical coordination of different building systems between disciplines which must take place for proper and efficient installation in the field. On any given construction project, multiple building systems which are installed by different subcontractors must be combined, managed and reconciled to prevent physical conflicts. Example *building trades* which must be coordinated on a typical project include: structural steel, mechanical systems ductwork and piping, lighting and power systems, plumbing systems, fire protection systems, and underground utilities. On large commercial construction projects,

Building Information Modeling (BIM) has made a significant impact on this process. In fact, 3D clash detection is just one of many preconstruction services which many owners now expect of construction managers. Tools such as Autodesk's Navisworks have improved the visualization, speed and accuracy of coordinating different disciplines and participating subcontractors and have been shown to be significantly more efficient than traditional means.

However, this process is not without its challenges. During BIM-assisted coordination, subcontractors will provide their *3D trade model* representing the physical geometries and dimensional information which encompass the building systems that they are responsible for installing. However, this process is not without its challenges. Different trades will submit several versions of their model representing different zones and levels of the coordinated building project at different points in time. On large projects, construction managers must sometimes track, document, and coordinate hundreds of model files per trade in an accelerated time period.

While the practice is aided by advancements in server technologies, the evaluation of these files and their changes provides a wealth of information to construction managers that may ordinarily go unnoticed. For example, by determining the number of changes to 3D modeled objects which occur within each version of a file, construction managers can begin to assess their subcontractors' model updating productivity. Once this information is available for a large sample of projects, predictions can be made regarding the length of the coordination process and which trades may be most influential in decreasing the estimated project duration.

This paper documents a case study of a sample BIM-assisted projects where such data was tracked and analyzed. It makes inferences about the kinds of information found in trade model files which may be useful to construction managers and evaluates the usefulness of a proposed tool for automating that process.

LITERATURE REVIEW

Little research has been conducted on the modeling productivity of construction trades and its influence on the duration of the BIM-assisted coordination process. However, there have been several published case studies of coordination efforts using Virtual Design and Construction (VDC) which outline the benefits and challenges of these procedures. In particular, the MEP coordination of Healthcare projects, where an excess of complex systems must be routed in limited plenum space, has been the subject of several publications (Manning and Messner 2008, Khanzode et al. 2008 and Harty et al. 2010).

Moreover, diverse patterns of activity, used to manage digital model coordination, have been studied to understand relationships between Architecture and Engineering professions and their shared 3D models. The relationships and workflows which exist between different A/E professionals and the shared models are not static, as described by industry standards and documents (Whyte 2011). The modeling effort and corresponding impact on the model level of development (LoD) were analyzed by two project case studies (Leite et al. 2011). The results showed that more details did not necessarily require more modeling effort. In addition, LoD should be determined based on the intended purpose of the models (Leite et al. 2011).

In fact, segregated organizational structures are found to limit BIM collaboration (Dossick and Neff 2010).

OVERVIEW

Data for this study was provided by one of the largest international general contractors in North America, labeled GC to protect their confidentiality. Consequently, they have been one of the leading construction managers to begin fully integrating BIM in their business processes, with over $30 billion in BIM project experience. In fact, the integration of VDC services has become such an integral part of the company's culture, that they now offer a specific BIM career track and multitude of specialized training for their BIM management staff.

In the Spring of 2010, the University of Florida's Center for Advanced Construction Information Modeling (CACIM) partnered with GC in order to gain access to and insight from some of their historical BIM-assisted project data. One of the initial goals of the research was to evaluate current processes and make suggestions for how they might be enhanced, particularly in the processing and documentation of trade coordination. Through years of experience, the company has perfected their BIM coordination procedure significantly. However, they were interested in determining whether any patterns could be observed which might aid them in more efficiently managing trades and shortening the overall duration of trade coordination across all projects as well as gauging productivity during model updating for coordination purposes.

The VDC Coordination Process. Like many construction managers, on a typical project, the GC divides the building into levels and zones and then creates a coordination schedule during preconstruction based on this organization. This is not unlike the traditional processes of the past which were conducted over a light table. However, the organization and visualization has been enhanced with the advent of modeling in 3D. Breaking the project into smaller zones and levels reduces the number of conflicts to a manageable amount and allows the sequence to follow the physical vertical construction of the building. In order to manage trades efficiently, weekly and sometimes biweekly meetings are scheduled to include all modeling parties in one location, where discussions take place as to how conflicts between different trades will be resolved. Prior to each coordination meeting, the latest version of each trade model for a scheduled level is combined into a master Navisworks Fileset (.nwf); clash batches between models are performed and decisions are documented as to how the clashes will be resolved before the subsequent meeting. The file is then archived to a Navisworks document set (.nwd) at the completion of the meeting and uploaded to the server for each party's reference. Each level may require several coordination meetings and the process is repeated until all major conflicts are resolved and the level is fully coordinated.

Between coordination meetings, each trade continuously uploads the latest versions of their model to the server, so that other trades can track their changes and resolve additional issues accordingly before the next meeting. The files uploaded are usually authored in some version of 3D AutoCAD (.dwgs). However, the Architecture and Structure disciplines are often authored in Autodesk Revit and then

converted to a Navisworks Cache file (.nwc). The upload of these different file versions may happen on a daily basis but sometimes even more frequently depending on the trade and the pro-activeness of the responsible modeling party. Sometimes new objects are modeled from version to version and other times; existing objects are merely shifted to harmonize with other trades' geometry.

As a result of this process, the number of files which are documented on the server for one project can range in the thousands. For quite some time, the GC had been interested in studying the information that exists within these files but had been unable to devote manpower or time to such a task due to the sheer volume of files.

Research Scope. As part of an on-going research study, CACIM was given access to a sample of completed BIM-assisted projects located on GC's VDC server. This server is the life-blood of VDC operations conducted by the company and is used to store information such as drawings, mark-ups, coordination documentation, models, and animations. For each trade model file which is uploaded during preconstruction, the server records information such as: the project code, file version, trade, level, zone, modeling author, file size and a time stamp. The information assists in the tedious task of file management and grants access to relevant real-time project information to all the parties involved in coordination.

METHODOLOGY

For this study, a sample building project which was coordinated through VDC conducted by GC was analyzed. First, all documented model files for the project were downloaded from the server, sorted and organized. They were grouped into folders by trade and then subfolders by level which contained each version of the files uploaded in chronological order. The researchers attempted to manually record file information into an MS Excel worksheet by opening each model file in Navisworks Manage and recording an inventory of the objects in each layer in order to track changes from version to version. The *layer* refers to a category that any object geometry is placed within, in the original 3D CAD authoring platform. For example, for the Mechanical DUCT trade files, the layers might include rectangular duct, round duct, air diffusers, fire dampers, and mechanical equipment which were further sorted into the supply, return, exhaust, and outside air systems. At first, 3D objects in each *layer* were manually selected from the Navisworks selection tree and the number shown in the properties palette toolbar was recorded to a spreadsheet for each file version. Forty files were analyzed in this way before it was determined that the volume of files was too large and time consuming to be analyzed manually.

Thus, work began on a plugin for Navisworks which could automate the process of opening files, counting the number of 3D objects modeled in each **layer** using a search set and then record the information in an Excel file. The tool was manually validated and the preliminary results of the tool were analyzed. In this paper, results are presented for one building trade on the project and recommendations are made as to how the procedure could be repeated across multiple building trades and projects. The intent of the tool is to be useful to a construction manager, so that they may begin to track object changes from file version to file

version, determine their trade subcontractors' productivity and better manage the modeling process.

RESULTS

Case Study. The example project used for this study is an eight story hospital coordinated in 2008. Table 1 describes the number of files uploaded to the server by each trade for each level of the building. Roughly 823 files were uploaded to the server over the course of the project because many trades submitted as many as 18 versions per level. Of that total, 152 files were actually utilized during coordination meetings, not including the master .nwf or .nwd files which were used to record clash batches. Thus, as much as 81% of the coordination process was conducted between trades through the use of file exchange on the VDC server outside of the group coordination meetings. The trades described in Table 1 are representative of the typical trades utilized on a project of this scale coordinated by the GC, but may vary slightly depending on the scale and type of project. A hospital project is referenced because of its complexity and the applicability of VDC at this scale. As shown in Table 1, there were a large number of files to be managed across different trades. There is no Level 4 shown in the table because on this project Levels 3 and 4 were identical. Thus, the coordination of Level 3 was used as reference for Level 4 due to time constraints.

Table 1. Number of File Versions Uploaded for a Typical Hospital Project

Trade	# of File Versions Per Building Level									
	L1	L2	L3	L5	L6	L7	LL	PH01	PH02	Total
ARC	18	18	18	18	18	18	9	9	9	135
ARCP	14	14	15	17	15	15	14	6	7	117
ATC	4			1						5
DUCT	18	20	24	0	16	13	14	2	1	128
DOMV	19	8	12	9	5	8	5			66
FIRE	8	15	11	8	7	6	4	2	1	62
LGHT		5	9	3	3	2	4	1		27
LOMV			7	6	1	1				15
MEDGAS	4	4	17	6	6	8	5	1		51
PIPE	15	10	11	14	8	6	5			69
PLUMB	17	15	18	11	5	6	6			78
PNTU	4	1	3	1		4				13
POWR	9	7	18	5	4	7	6	1		57
Totals:	120	117	163	101	88	94	72	22	18	823

Manual Investigation. Through physical analysis of the first 40 files in the DUCT trade, a better understanding of how trades modeled and organized geometry was uncovered. For this particular trade, the CAD-DUCT layers were first broken up similar to the different mechanical dry systems such as: Supply, Return, and Outside Air and then further into different object types within those layers such as: rectangular duct, round duct, and equipment. The layer organization gave insight into which object geometry changed most frequently from version to version and what areas proved to be most challenging to model for that trade. It should also be noted that the accuracy and efficiency of how different trades named and organized their

layers within their models was also found to be an indicator of their modeling proficiency. However, it was not analyzed as part of this study.

During analysis, some discrepancies were uncovered in how Navisworks itemizes objects in the selection tree. When imported into Navisworks, some 3D objects from CAD-DUCT were further subdivided into subparts representing individual faces of object geometry. For example, a segment of rectangular duct may show in the selection tree as one object with each rectangular face of the object denoted as children beneath it. The researchers were interested in the major object geometry being modeled, not the individual faces which make up an object. Thus, a custom search set was created which excluded the "sub-entity" objects and the layer name from being double counted in the final number. Figure 1 describes the search set in Navisworks which was created to help select objects to be counted. This search was utilized in the creation of the automated tool. Additionally, there were also a great number of layers which referenced 2D elements such as dimensions and annotations which were not representative of 3D geometry. For this study, all layers which included 2D in their name were removed from the final spreadsheet.
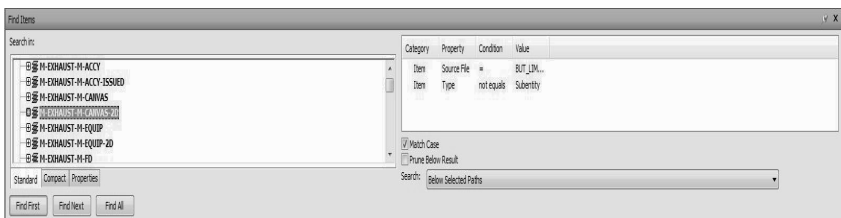
Figure 1. An example search set that was used to select objects in each layer

The Tool. The tool which was created to automate this process utilizes Navisworks API, which allows advanced users to code plugins that add functionality to existing Navisworks functions. The first part of the tool is a plugin to count the number of the actual model objects in a single model file, while filtering out the supporting entities including "File", "Layer" and "Sub-entity" as described previously. The second part of this tool uses the Navisworks Automation API, which triggers Navisworks functions (the single file counter in this case) over a bunch of model files without the user actually opening each model file.

For the first model file encountered by the plugin in the designated folder, the tool will record and write all the layer names and the number of objects in the layers to an MS Excel output file. For each subsequent model file, a new column is created in the same MS Excel worksheet. For each layer, the plugin will try to find a layer with the same name. If a matching layer is found, the number of objects is recorded in the same row using a new column. If the layer name is not recognized from before, a new row is added to the spreadsheet in the new column.

The object changes from version to version for a sample of layers from the Duct trade's Level 1 which were calculated by the tool are shown in Table 2. The tool was validated by checking its resulting output against the manual query results which were conducted for the first 40 files. Both the manual and automated results matched, thus validating the tool. Based on a comparison of results from manual

investigation, the tool proved to save significant amounts of time and reduce potential data entry error. Manual analysis of the first 40 files was estimated to take roughly five hours, whereas the tool was able to complete analysis of an entire trade in under10 minutes.

Table 2. Example Output for the Duct Trade on a Hospital Project

Layer Version	M-ACCY-A	M-ACCY-	M-Canvas	M-EQUIP	M-FD	M-FD-Issued	M-Flex	M-Flex-Issued	M-GRD-Diffusers	M-RECT
V0001	5	1	2	1	3	2		1	1	207
V0002	3		2	1	3	2				118
V0003	3		2	1	3	2				117
V0004	22	1	3	3	3	2	7	1	14	225
V0005	22	1	3	3	3	2	7	1	14	226
V0006	22	1	3	3	3	2	7	1	14	227
V0007	22	1	3	3	3	2	7	1	17	235
V0008	22	1	3	3	3	2	7	1	17	235
V0009	26	1	3	3	3	2	7	1	17	235
V0010	22	1	3	3	3	2	7	1	17	236
V0011	22	1	1	3	3	2	7	1	18	239

Duct Trade Productivity Results. In order to illustrate the potential applications for this tool, a sample analysis was conducted based on the findings uncovered from the Duct Trade. For example, Figure 2 shows the objects changes from version to version for Level 1's supply air layers. The different object layers are indicated in different colors to show which layers require more objects and which change most frequently. On this level, after the fourth version was uploaded (one month from the first version), the object number changes became stable; indicating that all major modeling by the Duct trade was completed within the first month. Therefore, most of the changes which took place after that time involved shifting or moving objects which were already previously modeled.
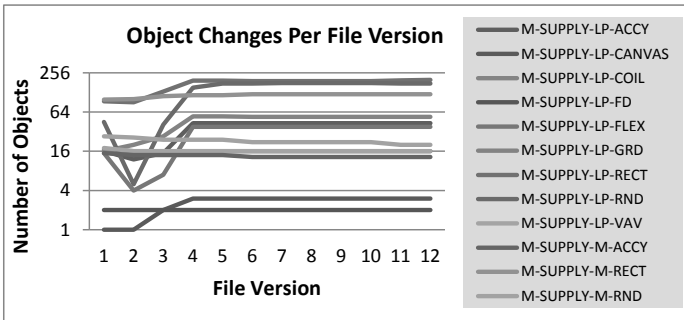
Figure 2. Number of object changes per version in Supply Air layers of level 1

This kind of information applied across each building trade is very useful to construction managers. If a similar pattern were uncovered in other trades models and across other projects, the GC would a have a realistic estimate for the time that it takes different trades to complete modeling and the time required to actually move

and coordinate with other trades. For example, based on the area (SF) of a level being coordinated and the time lapse between when different file versions were uploaded, the GC could start making predictions regarding the number of days per SF that each trade needs to adequately complete the modeling process. Additionally, if it is found that certain layers or object categories require much more labor than others; those object categories can be given priority during the coordination process.

CONCLUSIONS

Preliminary results from the automation tool proved to be very promising for the GC. Knowledge of different trades' modeling productivity may encourage healthy competition between subcontractors and allow GCs to evaluate future business relationships with those subcontractors more effectively. But at a minimum this tool presents an additional means for construction managers to better manage their subcontractors. It gives insight into the intricacies of the trade modeling process and provides a wealth of information which can be tracked across BIM-assisted projects.

This study focused on the modeling changes which took place in the Duct trade on one project. Future studies should compare the productivity between different trades within one project or one trade across several building projects in order to gain useful insight about the GC's coordination process as a whole.

REFERENCES

Dossick, C. and Neff, G. (2010). "Organizational Divisions in BIM-Enabled Commercial Construction," *Journal of Construction Engineering and Management* , 136(4), 459-467.

Hardy, C. and Throssell, D. (2010). "Implementing Building Information Modeling: A Case Study of the Barts and the London Hospitals," Proceedings from the *2010 International Conference on Computing in Civil and Building Engineering.*

Khanzode, A., Fischer, M. and Reed, D (2008). "Benefits and Lessons Learned of Implementing Building Virtual Design and Construction (VDC) Technologies for Coordination of Mechanical, Electrical, and Plumbing (MEP) Systems on a Large Healthcare Project," *ITCON*, 13, 324-342.

Leite, F., Aakcamete, A., Akinci, B., Atasoy, G. and Kiziltas, S. (2011). "Analysis of Modeling Effort and Impact of Different Levels of Detail in Building Information Models," *Automation in Construction* , 20, 601-609.

Manning, R. and Messner, J. (2008). "Case Studies in BIM Implementation for Programming of Healthcare Facilities" *ITCON*, 13, 446-457.

Whyte, J. (2011). "Managing Digital Coordination of Design: Emerging Hybrid Practices in an Institutionalized Project Setting," *Engineering Project Organization Journal* (September 2011) 1, 59-168.

GIS Assessment of the Emergency Medical Response Infrastructure

De-Ching Huang[1], Hsiao-Hsuan Liu[2], Albert Y. Chen[2], and Wei-Zen Sun[3]

[1]Fire Department, New Taipei City Government, No.15, Sec. 2, Nanya S. Rd., Banqiao Dist., New Taipei City 22060, Taiwan; PH +886-2-89519119; FAX +886-2-22514761; email: commissioner@mail.tpf.gov.tw
[2]Department of Civil Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617, Taiwan; PH +886-2-33664255; FAX +886-2-23639990; email: {b97501107, albertchen}@ntu.edu.tw
[3]Department of Anesthesiology, National Taiwan University Hospital, No.7, Zhongshan S. Rd., Zhongzheng Dist., Taipei City 10002, Taiwan; PH +886-2-23123456-62158; FAX +886-2-23415736; email: wzsun@ntu.edu.tw

ABSTRACT

The effectiveness of emergency medical service (EMS) depends on the existing physical infrastructure and the allocation of medical resources. EMS is a challenging task due to the spatial distribution of the population and the geographical layout in the urban region. The management of emergency medical units (EMUs) and hospitals should be assessed to provide an efficient service. The objective of this research effort is to assess the current EMS of New Taipei City using geospatial analyses such as facilities' service area. In this study, spatial statistics for EMUs and hospitals are conducted. In addition, regions that cannot be reached within 10 minutes under pre-assumed conditions by EMUs are discovered, and interpolation of the travel interval is conducted and investigated. The geospatial arrangement of EMUs, and the upgrade or new constructions of hospitals are suggested to strengthen the efficiency of the emergency medical response. Future work is directed towards more advanced and complex spatial and temporal analyses to better assess the city's medical infrastructure.

INTRODUCTION

Emergency Medical Service (EMS) includes on-site medical treatment of injured or ill patients and the transport of patients from the incident to a hospital in emergency situations (Laws and Regulations DB of ROC, 2007). High quality EMS secures life and health of the patients in such cases. For severe cases, the efficiency of EMS arrangements is critical to the survival rate. New Taipei City, the city in Taiwan with very dense population, has great demand for EMS (Liu and Chen, 2012; Sun et al., 2012). However, due to the vast extent of its territory, unique geospatial layout, and uneven distribution of population and medical resources, it is challenging to efficiently dispatch emergency medical units (EMUs) in New Taipei City. There are

also challenges in the present Emergency Medical Service System (EMSS) of New Taipei City, such as the allocation of its resources and the systematic planning and execution of its services. These factors suggest seeking of an effective way to integrate the EMSS of the City.

In New Taipei City, EMUs fall under the command of the city's Fire Department. Under the Fire Department, there are sixty-seven EMUs, which are in charge of different areas in the city. On the other hand, there are fifty-five hospital candidates located in New Taipei City and the other three cities near New Taipei City. Throughout an emergency medical response, spatial factors play important roles in its success (Ko, 2008). For major trauma patients, the travel time between the EMU and the scene, and the travel time between the scene and the hospital are both decisive of the mortality. As a result, spatial factors of emergency medical actions should be taken into considerations.

LITERATURE REVIEW

There are studies in the field of emergency medical dispatch, and the followings are those that relates to the study in this paper. Warden (2007) concluded that Geographic Information System (GIS) is a good method for analyzing EMS data due to its addition of spatial components to traditional approaches. In recent years, GIS has assisted the EMS domain in many ways. Ong et al. (2008) had used GIS with a national cardiac arrest database to derive the geographical distribution pattern of pre-hospital cardiac arrest in Singapore. Pathak et al. (2011) used GIS to analyze the cluster of non-transported and transported cardiac decedents in terms of distances to the closest hospitals. GIS can also simulate the choice of the transportation mode in EMS. Lerner et al. (1999) applied GIS and historical transport data to create a reference model that helps to outline the region where air or ground transport can minimize out-of-hospital time in emergency medical incidents. For assessing ambulance response performance, Peters and Hall (1999) developed a GIS framework and applied it to three communities in Southern Ontario, Canada. Peleg and Pliskin (2004) presented a GIS simulation model suggesting that a dynamic load-responsive ambulance deployment can make EMS more effective.

In terms of the use of GIS to health care services, Higgs (2004) pointed out that GIS has the potential to help identify less accessible areas and show the changes of facility re-locations. Henderson and Mason (2004) combined computer-aided dispatch (CAD) database, GIS visualization, and simulation to make a better ambulance service planning. Lim et al. (2011) found that adopting an appropriate ambulance dispatch policy with a GIS can provide faster response time. Yin and Mu (2012) applied a modular capacitated maximal covering location problem with GIS and optimization software to optimally allocate ambulances for EMS in the State of Georgia. Bailey et al. (2011) provided several candidate scenarios helping improve the emergency referral system in Ethiopia and used GIS to make the assessment. It is pointed out that GIS is helpful in identifying the priority for facility upgrading. For the GIS analysis in Taiwan, Ko (2008) applied GIS to analyze the EMS distribution and characters of cross-district EMS transport for major trauma patients in northern Taiwan. Ko pointed out that utilizing GIS may show the spatial characters of these

cross-district transports. Furthermore, spatial analysis is a good way to search the region of concern and the priority for EMS reconstruction.

The aforementioned works show that GIS can help analyze and identify problems of existing EMS. However, service areas and interpolation representations of both the EMUs and the hospitals have not been thoroughly conducted and discussed, based on the best knowledge of the authors.

OBJECTIVE

The objective of this work is to assess the medical infrastructure of New Taipei City. The main facilities considered are the location of EMUs and the hospitals. The first goal is to find out whether the current allocation and distribution of EMUs and the hospitals are proper. The second is to assess selection criteria of hospitals. Further is to analyze the medical infrastructure with spatial analyses. In addition, temporal analysis of the infrastructure is also considered important. The authors aim to discover the dispatch inefficiencies and black regions - places non-reachable in a reasonable amount of time for emergency transport - of EMS in New Taipei City.

APPROACH

In order to assess the emergency medical infrastructure in New Taipei City, GIS has been used to integrate all the related data. The analyses to discover the spatial characteristics of EMS include GIS spatial representation, facilities' service area, and spatial interpolation. The following subsections discuss these methods. The findings will be discussed in the Discussion section.

GIS Representation. The road network data and files used in this study are provided by the Institute of Transportation, Minister of Transportation and Communications in Taiwan. Data and files include the outlines for all the cities, districts, and towns in Taiwan. The EMS database provided by the Fire Department of New Taipei City Government is also used in this study. The EMU dispatched and the hospital selection of each incident has been recorded in the database.

From the EMS database, the number of total counts each EMU being dispatched in 2010 is obtained, as well as the number of total counts each hospital being selected. Those numbers have been ported to the GIS software, ESRI ArcGIS Desktop 10. In order to better visualize the data, the authors input the x-y coordinate of each EMU and hospital. The total counts of each EMU being dispatched in 2010 are displayed in Figure 1(a). On the other hand, the total counts of each hospital being selected in 2010 are displayed in Figure 1(b). The background map of Figure 1 displays the districts and the population distribution in New Taipei City. The darker the color of the district, the greater is the population of the district.

Service Area. The service area of both the EMUs and the hospitals are analyzed in this section. The analysis has been set to limit the network length to 10 km with the assumption that the average travel speed of ambulances is 60 km/h and that both the travel time between the units to the scene and the scene to the hospitals

should not be greater than 10 minutes. Figure 2(a) displays the 10 km service area of all the 67 EMUs in New Taipei City. On the other hand, the 10 km service area of all the 55 hospitals is shown in Figure 2(b).
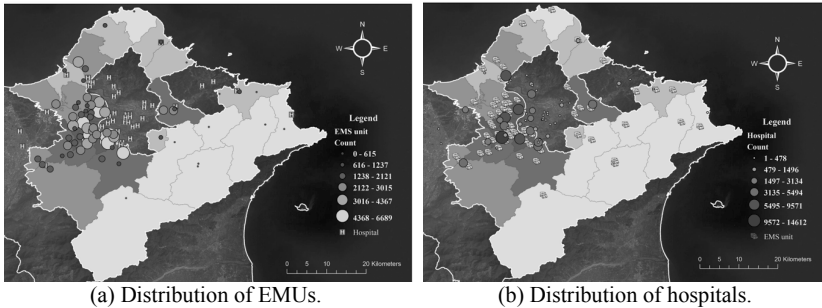
(a) Distribution of EMUs. (b) Distribution of hospitals.
Figure 1. Distribution of EMUs and Hospitals and Their Service Counts.
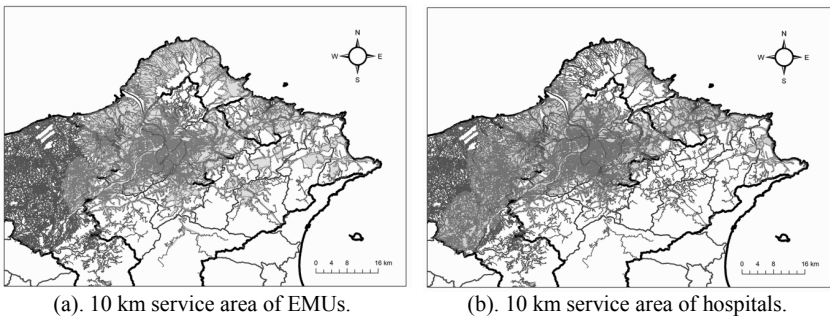
(a). 10 km service area of EMUs. (b). 10 km service area of hospitals.
Figure 2. Service Area of EMUs and Hospitals.

Transported Incident Analysis. Ten hospitals have been chosen to analyze the transported incidents. The incidents that have been sent to these ten hospitals are investigated in Figure 3. A circle marker in Figure 3(a) represents a hospital and is one of the ending nodes of a link. The other ends of the links are the locations of the incidents. If the colors of the lines are the same, it means that those lines intercept at the same circle; in other words, the incidents are sent to the same hospital. Figure 3(b) depicts the monthly counts of these ten hospitals.

Non-Transported Incident Analysis. It is possible that the EMU has arrived at the scene but does not transport the patient to any hospital. This type of EMS incidents is called the non-transported incident. There are many reasons that may lead to it, such as the patients' refusal to be sent to hospitals, and the patients' being dead before transporting to hospitals. The non-transported incidents may waste the resource of the EMS system. It is better to reduce the number of the non-transported incidents. Figure 4(a) displays the monthly counts of the non-transported incident for ten EMUs which have been dispatched to the non-transported incidents most. The

response time plus on-scene time is the time a non-transported incident waste. Figure 4(b) indicates the average response time plus on-scene time for the same ten EMUs.
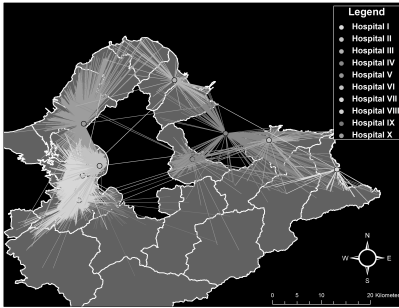
(a). Spatial Relationship of Hospitals and Incidents.
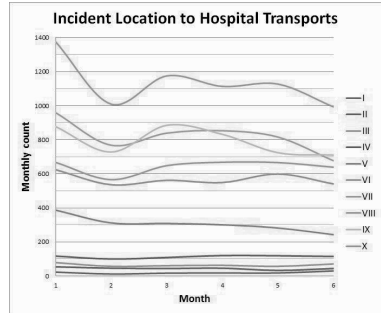
(b). Temporal Relationship of Hospitals and Incidents.

Figure 3. Transported Incidents for 10 Hospitals.
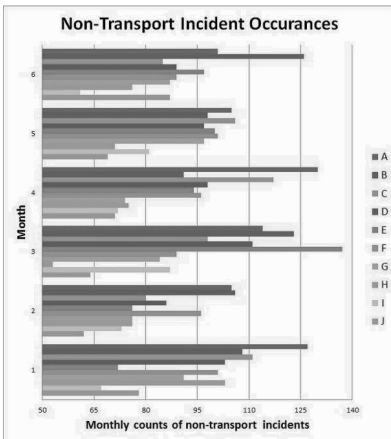
(a) Incident Occurrences
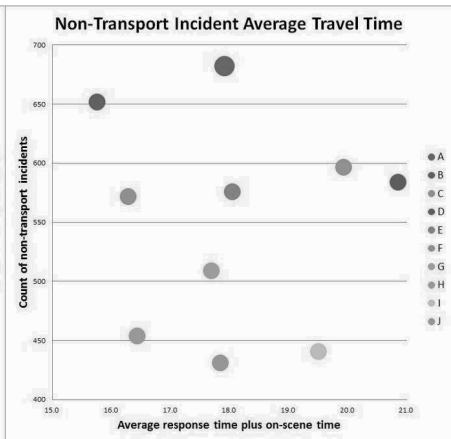
(b) Average Travel Time

Figure 4. Non-Transported Incidents for 10 EMUs.

Spatial Distribution. Contour maps are visual representations to investigate spatial distribution of points with specific attributes. Kriging is selected in this study to interpolate the data between points to produce this contour representation. Kriging is a statistical method that helps to interpolate the weight of unknown surface, with weights declining according to the distance between the points (Harvard School of Public Health, 2013). It assumes that the data points that are closer to each other are more similar than points that are far away from each other. Kriging is conducted to visualize the combination of spatial and temporal characteristics of the response time and the transport time of EMS incidents. Figure 5(a) shows the Kriging result for the response time, and Figure 5(b) shows the Kriging result for the transport time.
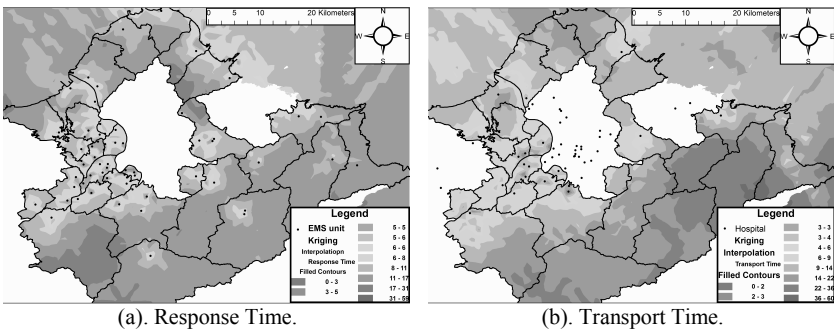
(a). Response Time.   (b). Transport Time.
Figure 5. Spatial Interpolation of Emergency Incidents.

DISCUSSION

In previous sections, results of various analyses are presented. This section discusses the results regarding the assessment of the current medical infrastructure.

Figure 1(a) and Figure 2(a) show that the allocation and distribution of EMUs are proper. The density of the units is in accordance with the population. The areas that are not covered by the 10 km service area are mostly in the mountains where population is less. Figure 1(b) and Figure 2(b) show the allocation and distribution of the hospitals. Due to limitation of medical resources, hospitals and EMUs tend to be located at areas with larger population. However, there are six districts in the city that are almost not covered by the 10 minute service area. In four of these six districts, most of the territories are mountains and the populations are below 8,000. These might be reasons why these districts are not in the service area of the hospitals. However, there are famous tourist spots in these 4 districts. Lacking hospitals that can be reached in 10 minutes threatens the survival rate of major trauma patients during holidays. The populations of the other two districts are 12,700 and 23,263. It is improper that they are not included in the 10 minute service area of any hospital.

The areas that are not covered by the service areas of both EMUs and hospitals cannot be reached in 10 minutes by both the units and the hospitals. It may take more than 20 minutes for the unit to get to the scene and deliver the patient to a hospital. The 10 minutes travel time might be too long for major trauma patients. These patients need to be stabilized within the golden period of 4 to 6 minutes. In certain cases, brain cells of the patient are damaged and those cells cannot be recovered. As a result, to have short response time is very important.

Figure 3(a) shows that aside from distance, there may be some other factors that affect the choice of the hospitals. After reviewing the EMS database, the authors found that those factors include the condition of the patients and special requests from patients' family members. The latter is the human factor that is difficult to analyze and change. However, the other factor reveals that the medical resource and capability of a certain hospital affect the choice of hospital selection.

Figure 4 can help understand to what extent the non-transported incidents influence the EMUs. Take unit E as an example. Its average response time plus on-scene time is about 18 minutes. In June, unit E faced about 100 non-transport

incidents. As a result, EMS spent about 30 hours dealing with non-transport incidents. It means that an hour per day had been wasted in June for EMS unit E.

The combination of spatial and temporal characteristics by using Kriging helps examine the performance of current EMS in New Taipei City. The result in Figure 5(a) is reasonable. The response time for most of the areas near EMUs is less than 8 minutes, and the populous districts are mostly less than 5 minutes. It shows that the current EMS infrastructure perform well in response time. However, there are still some problems. The response time of areas near the EMU located in the easternmost district is greater than 11 minutes, which is dangerous for major trauma patients. After checking the incidents taking place in that district, we realize that those incidents are located near the east coast of the district (tourist area) while the EMU is located in the middle mountainous area of the district. The location of the EMU seems to be improper and should be reconsidered.

CONCLUSION

The allocation and the distribution of EMUs and hospitals severely affect the EMS system. The majority of the population should be covered in the service area, and a good emergency medical dispatch pattern may enable a more efficient EMS arrangement. The emergency medical infrastructure in New Taipei City is in general proper. The city has more dense medical services at regions with dense population. However, response time and transport time for regions with less EMUs take significantly longer and this poses great risk to trauma patients. The geospatial arrangement of EMUs and upgrade or new constructions of hospitals in this region are suggested to strengthen the efficiency of emergency medical response. Additional work is directed towards more advanced and complex spatial and temporal analyses to better assess the city's medical infrastructure.

ACKNOWLEDGEMENTS

REFERENCES

Bailey, P. E., Keyes, E. B., Parker, C., Abdullah, M., Kebede, H., Freedman, L. (2011) "Using a GIS to model interventions to strengthen the emergency referral system for maternal and newborn health in Ethiopia." *International Journal of Gynecology and Obstetrics*, vol. 115, pp. 300–309.

Harvard School of Public Health. (2013). "Geographic Information Systems (GIS) in Public Health Research – Kriging." (Jan. 14, 2013). http://www.hsph.harvard.edu/research/gis/arcgis-tips/kriging/>

Henderson, S. G., and Mason, A. J. (2004). "Ambulance service planning: simulation and data visualization." *ORHC*, vol. 70, pp. 77–102.

Higgs, G. A. (2004). "A literature review of the use of GIS-based measures of access to health care services." *Health Services & Outcomes Research Methodology*, vol. 5, pp. 119–134.

Ko, C. I. (2008). "Spatial Distribution and Influence Factors of Cross-District Transports among major trauma in emergency medical services system." *Master Thesis, National Taiwan University*.

Laws & Regulation Database of The Republic of China (Laws and Regulations DB of ROC). (2007). "Emergency medical services act." (Jan. 14, 2013). http://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020045>

Lerner, E. B., Billitier, A. J. IV, Sikora, J., and Moscati, R. M. (1999). "Use of a geographic information system to determine appropriate means of trauma patient transport." *Academic Emergency Medicine*, vol. 6, pp. 1127–1133.

Lim, C. S., Mamat, R., Braunl, T. (2011). "Impact of ambulance dispatch policies on performance of emergency medical services." *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 624–632.

Liu, H.-H. and Chen, A. Y. (2012). "Emergency Medical Dispatch: A Case Study of New Taipei City." Presented at 12th International Conference on Construction Application of Virtual Reality (CONVR 2012). Taipei, Taiwan.

Ong, M. E., Tan, E. H., Yan, X., Anushia, P., Lim, S. H., Leong, B. S., Ong, V. Y., Tiah, L., Yap, S., Overton, J., and Anantharaman, V. (2008). "An observational study describing the geographic-time distribution of cardiac arrests in Singapore: What is the utility of geographic information systems for planning public access defibrillation?" *Resuscitation*, vol. 76, pp. 388–396.

Pathak, E. B., Reader, S., Tanner, J. P., Casper, M. L. (2011). "Spatial clustering of non-transported cardiac decedents: the results of a point pattern analysis and an inquiry into social environmental correlates." *International Journal of Health Geographics*, 10:46.

Peleg, K. and Pliskin, J. S. (2004). "A geographic information system simulation model of EMS: reducing ambulance response time." *The American Journal of Emergency Medicine*, vol. 22, pp. 164–170.

Peters, J. and Hall, G. B. (1999). "Assessment of ambulance response performance using a geographic information system." *Social Sci. & Med.*, vol. 49, pp. 1551–1566.

Sun, W. J., Ma, H. M., Yu, J. Y., and Chen, A. Y. (2012). "Applying enterprise management method – using balanced scorecard for emergency medical management." *Fire Department of New Taipei City Government*.

Warden, C. R., Daya, M., and LeGrady, L. A. (2007). "Using geographic information systems to evaluate cardiac arrest survival." *Prehospital Emergency Care*, vol. 11, pp. 19–24.

Yin, P., Mu, L. (2012). "Modular capacitated maximal covering location problem for the optimal siting of emergency vehicles." *Applied Geography*, vol. 34, pp. 247–254.

Integrating BIMserver and OpenStudio for Energy Efficient Building

Nan Yu[1], Yufei Jiang[1], Lannan Luo[1], Sanghoon Lee[2], Abdou Jallow[2], Dinghao Wu[1], John I. Messner[2], Robert M. Leicht[2], and John Yen[1]

[1]College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802. Email: {nuy106, yzj107, lzl144, dwu, jyen}@ist.psu.edu
[2]Department of Architectural Engineering, The Pennsylvania State University, University Park, PA 16802. Email: {Sanghoon.Lee, ajallow, jmessner, rmleicht}@engr.psu.edu

ABSTRACT

Energy Efficient Building (EEB) design requires many simulation tools to support making decisions for optimized building solutions, resulting in frequent interactions between computational tools. Building Information Modeling (BIM) server platforms can support the storage, maintenance, and query of IFC-based building information models. However, the lack of a unified interface to support information exchange and interoperability among different building design and simulation tools has become a bottleneck of the EEB design process. This paper describes an integrated approach at the data level to combining BIMserver and OpenStudio to build a unified EEB data exchange model. Our first step is to build an information exchange bridge between BIMserver and OpenStudio, which will enable different design and simulation tools that are connected to either of them to interoperate and exchange needed data. In the paper, we also discuss the challenges of the seamless integration due to the dependency on both BIMserver and OpenStudio. The integrated approach, which organizes the data flow in a unified model, enabling effective exchange of data, is currently in a beta-testing phase.

Keywords: BIMserver, OpenStudio, Energy Efficient Building (EEB), Data Exchange Model

INTRODUCTION

Energy consumption has gained much attention recently. In the Architecture, Engineering, Construction and Facilities Management (AEC/FM) field, energy efficient building design is becoming more critical, especially as it relates to energy retrofit projects. In the process of energy efficient building design, decision-making in the very early stages might significantly influence the energy consumption (Pollock et al., 2009). The decision-making process should be built upon a channel, which connects the computational representation of a building's energy elements and the corresponding economic considerations (Jones et al., 2010). Energy modeling is

such a channel providing designers with an outlook of potential energy consumption of varieties of designs prior to constructing the building (Fleming et al., 2012).

During the building design lifecycle, Energy Efficient Building (EEB) design depends on the collaboration project participants using a variety of simulation tools to make decisions for the optimized building solutions. Currently, different simulation tools running in different "energy simulation views" (Bazjanac, 2008) determine the varieties of data sets and data formats (Bazjanac & Kiviniemi, 2007). It is necessary to gather all simulation views into an integrated whole-building simulation methodology with the intent of exchanging data seamlessly (Guglielmetti et al., 2011).

Some Building Information Modeling (BIM) server platforms, such as 'BIMserver' implemented by bimserver.org (Beetz el al., 2010) support the storage, maintenance, and query of Industry Foundation Classes (IFC) based building information models. OpenStudio, an interface to support whole building energy-related modeling and simulations, has another set of building energy modeling (BEM) representations (Weaver et al., 2012). A disconnect between different models prevents architects, engineers, and researchers from easily conducting integrated whole-building energy analysis (Guglielmetti et al., 2011). Accordingly, this paper explores an integrated approach to leverage BIMserver and OpenStudio to enable open data exchange and interoperability among different building design and simulation tools. With the integrated approach, the inherent data inconsistency and mapping problems could also be solved.

BACKGROUND

BIM for energy efficient building design. The AEC/FM industry shows increasing interest in adopting information technology in building designs (Bazjanac & Kiviniemi, 2007). BIM acts as a bridge between the industry and information technology (Eastman et al., 2008), which makes the entire building lifecycle more efficient and effective. BIM, an interoperable data model, can enable bi-directional data service for various simulation tools in the AEC/FM industry projects, which means the import and export of relevant data must be compatible with other tools (Bazjanac, 2007; O'Donnell et al., 2011).

Current simulation tools for building energy analysis. Building energy performance simulation leverages computer-based building energy analysis to quantitatively validate the correctness of decisions on building design and operations (Bazjanac et al., 2011). High-performance buildings require an integrated analysis, including whole-building energy, daylighting, airflow, among others. Traditionally, different simulation tools focus on their own domains. For instance, CONTAM is a multi-zone airflow and contaminant transport analysis software (Walton & Dols, 2010), EnergyPlus performs whole-building energy analysis, and Radiance is used for daylight and electric lighting simulation (Guglielmetti & Scheib, 2012). Typically, designers leveraging these tools use a point-to-point data exchange model (O'Donnell et al., 2011), which is often very complicated and inefficient. In order to facilitate information flow from an architectural design model to an energy model, Bazjanac and Kiviniemi (2007) developed a tool called the Geometry Simplification Tool (GST)

that simplifies valid IFC geometric model and extract construction properties.

Building information representations. One of the essential barriers preventing seamless data exchange is the different building information representations. Each representation offers a range of proprietary file formats used in modeling and simulation tools; EnergyPlus input data file (IDF), Green Building XML (gbXML) and DXF are some of the most common data representations and input file formats used in energy-related simulation tools. On the other hand, Industry Foundation Classes (IFC) is the most mature open standard to represent building information and some BIM tools including the BIMserver support the storage, maintenance and query of the IFC-based BIM models. Due to the difference between the information representations, the transformation among them introduces human intervention, which also makes the process inefficient and error prone. Therefore, a highly integrated and easily shared information exchange model is required.

THE INTEGRATED DATA EXCHANGE MODEL

The commercially available BIM authoring applications do not include the HVAC system or operation schedules within the IFC export. Therefore, a link between building information exported in IFC format from BIMserver and BEM is required for an integrated data exchange model to support the whole simulation process.

Some initial work to investigate information exchange between BIM tools and simulation tools was performed using airflow analysis and to develop the integrated data exchange model to support the interoperability. Figure 1 shows an airflow analysis model development workflow using BIM (Jiang et al., 2013). The horizontal level stands for the simulation process inside CONTAM, including zoning, specifying flow path, connecting HVAC to zones and running the simulation. It is of critical importance that we build a tool to collect building information specified by the requirements and connect CONTAM with BIMserver. During the information collection process, a subset of building geometric information would be extracted via a query function according to MVD (model-view-definitions). The output of the tool is a subset of input data that CONTAM understands.
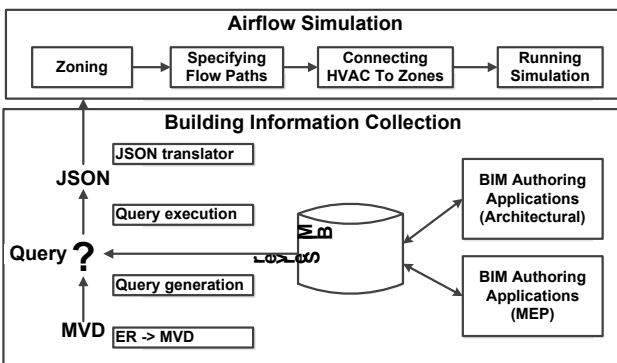
Figure 1. A Representative Workflow: Airflow Analysis Model Development Using BIM

Integration through OpenStudio

The workflow in Figure 1 provides a feasible method for airflow analysis through direct information extraction from an IFC model into CONTAM. However, as mentioned in the background section, building information has various representations and building energy modeling usually works in different "energy simulation views". A point-to-point communication channel needs to be established for each direct data exchange. The complexity of the point-to-point information exchange model would increase dramatically as the number of simulation tools increases. Thus, the AEC/FM industry experts and IT professionals are pursuing integrated models to eliminate the need for distinct data models and management services at the individual tool level. Jiang et al. (2012) propose that the integration of BIMserver and OpenStudio might be a potential solution for information exchanges in the energy simulation domain. Based on this assumption, we have analyzed the feasibility to choose OpenStudio as a middleware interface.

OpenStudio is designed to establish an object-oriented framework for Building Energy Models (BEM), which is compatible with existing work (Guglielmetti, 2011). BIMserver is an object-oriented framework for building information and transforms all semi-structured information presented in IFC into building objects with corresponding attributes and relationships stored in memory. In view of the object-oriented characteristic, it is possible for data to flow from the BIMserver to OpenStudio and build an information exchange bridge for building simulations. Second, OpenStudio and EnergyPlus are bound together in the initial design (Ellis et al., 2008). Accordingly, EnergyPlus leverages OpenStudio to conduct whole-building energy analysis without extra effort. OpenStudio also supports Radiance to perform daylighting analysis besides conducting whole building energy analysis through EnergyPlus. It is desirable to adopt a single tool to facilitate analyses from different perspectives.

OpenStudio is designed to overcome the shortage of user-friendliness of EnergyPlus and Radiance. Lack of an intuitive graphical user interface (GUI) for the tools creates a high entry barrier which prevents novel users from using the tools (Weaver et al., 2012). Other software applications such as DesignBuilder, IES, and eQuest try to present a state-of-the-art GUI to users. However, some of these software applications are commercial software. Users are constrained by the GUI provided to make limited analyses. In contrast, OpenStudio is open-source, cross-platform and cross-language. Last but not least, OpenStudio provides a rapid development mode and open application programming interface (API), which makes it highly extensible and customizable. It is rather simple for developers to either build on existing applications or create completely new ones to conduct customized building energy analysis (Weaver et al., 2012). All of these aspects suggest OpenStudio as a suitable platform for initial targeting to support the data exchange needs of building energy modeling.

The proposed integrated data exchange model

Figure 2 shows the proposed framework of integrating BIMserver and OpenStudio to build an integrated EEB data exchange model. The first step is to build an information exchange bridge between BIMserver and OpenStudio, which enables different design and simulation tools to connect to either of them to interoperate and exchange needed data.
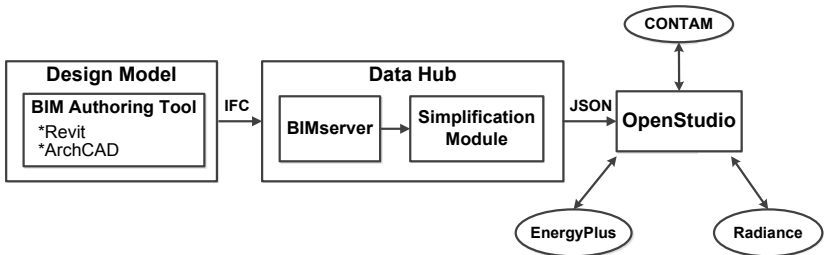
Figure 2.   The Proposed Integrated EEB Data Exchange Model

In the model, BIMserver is adopted as the implementation of a BIM server which takes charge of building information management and selective building information retrieval. It accepts IFC format exported from BIM authoring tools such as Autodesk® Revit® or other tools that support IFC exporting, as input data. The input IFC data contains overall and original building geometric information and is loaded into the memory of BIMserver.

However, OpenStudio requires a reduced and simplified data set instead of the complex and rich BIM data. Therefore, the exchange needs to extract the simplified geometry from memory and transmit it to OpenStudio, so a simplification module is shown inside the model. It automatically generates Java query code according to IFC specification. The Java code is inserted and run in the advanced query area in the GUI of BIMserver, which searches the memory and displays the query result in the console. The simplification module not only provides data set reduction and simplification, but also offers the function of data translation and data interpretation, which transforms the retrieved information into the appropriate data set and format to keep the inconsistency in different views. For example, most elements in a building are described under different local coordination systems, but the resulting set is required in the global coordination system. Therefore, the translation between various local coordination systems is part of the simplification module. The translation leverages the matrix transformation to relocate the origin and the directions of the axes. Furthermore, IFC specification is illustrated in an inheritance graph, where the bottom entities inherit the attributes and relationships from the top entities. In that case, the attributes and relationships of the bottom entities are interpreted by the derived data. Through reduction and simplification, data translation and interpretation, original data is transformed into the required data set available under different views. The above steps are to prepare the data structure.

Another function of the simplification module is data transformation, including the relationship mapping and format transformation. Data that flows into BIMserver is in IFC format, while the output format is supposed to be OpenStudio Model (OSM). These two different formats contain different sets of data properties. The mapping process from IFC to OSM is based on the properties analysis in different views. The format transformation is accomplished via JSON format. JSON (JavaScript Object Notation) is a lightweight open data-interchange format (http://www.json.org). All the data which passes through BIMserver is object-oriented so that it can organize the data as building objects and export the query result in JSON format. The data in the JSON format needs to be parsed and re-organized into OSM format, which is the required and accepted input format by OpenStudio. This can be implemented by a programming script in either Ruby or C#.

With the link between BIMserver and OpenStudio, the integrated model enables data exchange and the pipeline of whole-building energy. The pipeline refers to a chain of building information processing stages. For example, data processed in a design model is exported into the IFC format and taken by BIMserver, which selectively transforms the needed data for OpenStudio. Other tools linked with OpenStudio can run whole-building energy simulations to conduct energy analysis. In the data transformation from BIMserver to OpenStudio, the middleware interface can extract geometric points of building elements and simplified elements, such as walls with regular shapes. We also analyzed the concrete building examples in both IFC and OSM in expectation of creating transformation rules from IFC to OSM. With the integrated data exchange model, we are able to pipeline the data flow in a unified interface and thus enable effective exchange of data.

Challenges

Integrating simplification module into BIMserver. Currently the simplification module and BIMserver run in their own isolated workspaces. Users still need to paste the automatically generated Java query code from the simplification module into the GUI provided by BIMserver and then copy the query result back to the simplification module for exporting. It would be more efficient to eliminate human intervention completely by integrating the simplification module into BIMserver as a plug-in. BIMserver is an open-source project that allows modifying the source code and adding customized modules to serve for the integrated data exchange model.

Seamless information exchanges from simplification module to OpenStudio. The integration of the BIMserver and OpenStudio has a dependency on the JSON format which bridges the gap between IFC and OSM formats. Presently, OpenStudio does not support parsing JSON format directly but offers an open rapid development mode and API. As a result, the JSON format produced by the simplification module will rely on a script to transform JSON into OSM, which adds a new dependency on an extra external module. It is suggested that such a function should be embedded inside the framework of OpenStudio.

CONCLUSION

A high-performance building requires comprehensive whole-building energy analyses to optimize energy consumption. AEC/FM industry experts should leverage information technology and various simulation tools to conduct integrated analyses to assist with their decision-making in the whole building lifecycle. This paper proposes the concept of the unified EEB data exchange approach that integrates BIMserver and OpenStudio, and analyzes the feasibility of an integrated model. In the EEB data exchange model, BIMserver performs the information management and selective building geometry retrieval, and OpenStudio plays the role of middleware connecting other simulation tools. The unified model shows the potential of supporting domain experts to easily conduct various energy-related analyses without reproducing the same building information. The next step is to complete the beta test of the data exchange model. The prospect of the integrated model also drives us to overcome the challenges above and set up an enhanced EEB data hub encompassing both BIMserver and OpenStudio to achieve seamless data exchange.

ACKNOWLEDGEMENT

REFERENCES

Bazjanac, V. & Kiviniemi, A. (2007). "Reduction, simplification, translation and interpretation in the exchange of model data." *In Proceedings of the 24th Conference. Bringing ITC knowledge to work*: 163-168.

Bazjanac, V. (2008). "IFC BIM-based methodology for semiautomated building energy performance simulation." *In Proceedings of the 25th International Conference on Information Technology in Construction*, Santiago, CL: 292-299.

Bazjanac, V., Maile, T., O'Donnell, J., Rose, C., Mrazovic, N., (2011) "Data Enviroments and Processing in SemAutomated Simulation with EnergyPlus." *In: CIB W078-W102: 28th International Conference*. CIB, Sophia Antipolis, France.

Beetz, J., Berlo, L.-v., Laat, R.-d. and Helm, P.-v., (2010). "Bimserver.org – An Open Source IFC Model Server". *In Proceedings of the CIP W78 conference*, Cairo.

Eastman, C., Teicholz, P., Sacks, R. and Liston, K., (2008). "BIM Handbook", Wiley & Sons.

Ellis, P.G., Torcellini, P.A., Crawley, D.B.. (2008). "Energy Design Plug-in: An EnergyPlus Plug-in for SketchUp." *In Proceedings of the IBPSA-USA SimBuild Conference*, Berkeley, California.

Fleming, K., Long, N. & Swindler, A.. (2012). "The Building Component Library: an Online Repository to Facilitate Building Energy Model Creation." *In Proceedings of the 2012 ACEEE Summer Study on Energy Efficient Buildings*. Pacific Grove, Calif.

Guglielmetti, R., Macumber, D., Long, N. (2011) "OpenStudio: an open source integrated analysis platform." *In Proceedings of the 12th Conference of International Building Performance Simulation Association*, Sydney, Australia.

Guglielmetti, R., Scheib, J., (2012) "Challenges to Integrated Daylighting and Electric Lighting Simulation Methods in a Whole-building Energy Simulation Context." *In Proceedings of the 2012 Simbuild Conference*, August 2012, Madison, WI.

JSON.org. "Introducing JSON". available at *http://www.json.org/*.

Jiang, Y., Ming, J., Wu, D., Yen, J., Mitra, P., Messner, J.I., and Leicht, R. (2012). "BIM Server Requirements to Support the Energy Efficient Building Lifecycle", *In Proceedings of the 2012 ASCE International Conference on Computing in Civil Engineering*. Clearwater Beach, FL.

Jiang, Y., Ming, J., Wu, D., DeGraw, J., Lee, S., Jallow, A., Yen, J., Mitra, P., and Messner, J.I. (2013). "BIM Enabled Energy Efficient Building Analysis: Improving Software Interoperability in the AEC Community." *Manuscript in preparation*.

Jones, B., Bogus, S.M. (2010). "Decision Process for Energy Efficient Building Retrofits: The Owner's Perspective." August, Vol. 5, No. 3, pp. 131 - 146.

Kangaraj, G., Mahalingam, A. (2011). "Designing energy efficient commercial buildings - A systems framework. Energy Buildings." 43:2329–2343.

Klotz, L. (2011). "Cognitive biases in energy decisions during planning, design and construction of commercial buildings in the United States: An analytical framework and research needs." *Energy Efficiency*. 4:271–284.

O'Donnell, J., See, R., Rose, C., Maile, T., Bazjanac, V., Haves, P., (2011) "SimModel: A domain data model for whole building energy simulation.", *In: IBPSA Building Simulation*. Sydney, Australia.

Pollock, M., Roderick, Y., McEwan, D. & Wheatley, C. (2009). "Building simulation as an assisting tool in designing an energy efficient building: A case study." *In proceedings of 11th international conference of building simulation Glasgow*. pp.1191–1198.

Walton, G., Dols, W. (2010) "CONTAM User Guide and Program Documentation." http://www.bfrl.nist.gov/IAQanalysis/docs/CWHelp30.pdf, Last updated Dec. 25th, 2010.

Weaver, E., Long N., Fleming, K., Schott, M., Benne, K. and Hale, E. (2012) "Rapid Application Development with OpenStudio." *2012 ACEEE Summer Study*, Pacific Grove, California.

A Practice oriented BIM framework and workflows

M. Kassem[1] , N. Iqbal[2] and N. Dawood[1]

[1]Technology Futures Institute, School of Science and Engineering, Teesside
University, TS1 3BA, Middleborough, Borough Road;
PH (44) 0164 234 2494; email: M.kassem@tees.ac.uk
[2]BIM Academy, School of the Built and Natural Environment, Northumbria
University, NE1 8ST, Newcastle upon Tyne, Northumberland Road;
PH (44) 0191 227 4533; email: Nahim.iqbal@bimacademy.ac.uk

ABSTRACT

Various Building Information Modelling (BIM) frameworks and workflows
have been developed over the last few years. A BIM framework is a theoretical
structure explaining or simplifying complex aspects of the BIM domain by
identifying meaningful concepts and their interrelationship. BIM workflows are
structured information (e.g. process maps) intended for operational applications of
BIM concepts and tools. The majority of BIM frameworks and workflows, developed
in the past years, were either intended to build broad understanding and adoption at
industry level or specific for BIM usage in large enterprises. This study will illustrate
the development and application of a practice-oriented BIM framework that can be
used at project level.

INTRODUCTION

BIM is receiving an ever-increasing acceptance in the building industry as
organisations understand its potential in improving efficiency throughout the building
lifecycle. The key emphasis of BIM has been on interoperability and visualisation,
whereby consistent and accurate information can be communicated across the
lifecycle process (Eastman et al, 2011). However, BIM have been attributed
transformative capabilities within the Architecture, Engineering, Construction and
Operations (AECO) industry (Succar, 2009) triggering the need for BIM frameworks.
Indeed, there have been a number of BIM frameworks and workflows over the last
few years. Theoretical BIM frameworks, such as the ones developed by Succar
(Succar, 2009), aim to defining the general BIM requirements and domains of
knowledge in different fields (i.e. technology, process and policy). While these
theoretical frameworks cannot be utilized for practical implementation of BIM in
building projects, their domains of knowledge could be exploited to develop
practice-oriented BIM frameworks and workflows. Existing BIM workflows and
protocols (e.g USACE, 2010; GSA, 2010, Penn State, 2010; NIST, 2007; NY DDC,
2012; BCA Singapore, 2012; AGCA, 2007) are intended to facilitate BIM
implementation at either a broad industrial level or in specific enterprises and their
supply chains. As a result, there is still a need for the development of practice-
oriented BIM frameworks that could be applied at project level.

The overarching aim of this paper is to present a BIM framework and BIM workflows that can be used in building projects to increase the efficiency of processes and enhance the quality of information to all stakeholders involved in the project lifecycle. This work was undertaken as a collaborative effort between the industry and academia within a 'Knowledge Transfer Partnership' (KTP) scheme. A KTP is British government-supported scheme, which facilitates the interactions between a company base and an academic base, enabling businesses to use the research skills and qualities of academic institutions to address important business challenges. In a KTP scheme, both the strategic challenges affecting the company and a roadmap for addressing them are established at the outset of the project which typically lasts between one and four years. The remaining sections will: briefly introduce the UK BIM initiatives and design processes; explain the methodology used to develop the practice-oriented BIM framework; illustrate the main elements of the BIM framework, and demonstrate the testing of the framework in a 48 hour design competition challenge.

BIM INITIATIVES AND DESIGN PROCESSES IN THE UK

The British Standards Institution (BSI) has released 'BS 1192' standard that provides guidelines to support collaboration and BIM implementation. The BS 1192 establishes the methodology for managing the production, distribution and quality of construction information using a disciplined process for collaboration and a specified naming policy. The UK government has recently established a BIM task group who has established, in collaboration with the Construction Industry Council (CIC), 11 regional hubs to raise awareness of BIM and its benefits and facilitate the adoption of BIM processes and working methods in the UK construction industry. The Government's BIM Task Group in collaboration with the BSI has just published the third draft of 'PAS 1192-2:2012' which identifies the information requirements for the capital delivery phase of construction projects (BSI, 2012).

The process for organizing and managing the design phase of building projects in the UK is specified by The Royal Institute of British Architects (RIBA) and called RIBA 'Oultine Plan of Work' (figure 1). It is important to emphasize that although the RIBA Outline Plan is commonly illustrated as a start to end process, in reality it is of an iterative nature, whereby information at various stages is reviewed and modified depending on project requirements. Any BIM framework and workflow, including the processes for the different BIM workstreams at the different stages of the project lifecycle and the interoperability requirements, must consider this Outline Plan of Work. In turn, the description of the different work stages of the Outline Plan of Work has to be modified with changes that consider the BIM overlay. Indeed, these changes were just recently released by the RIBA in a document called 'BIM Overlay to the RIBA Outline Plan of Work' (RIBA, 2012) As part of the framework developed in this research, an alignment model between: 'BIM workstreams, technologies and interoperability' and the RIBA Outline Plan of Work, is produced.
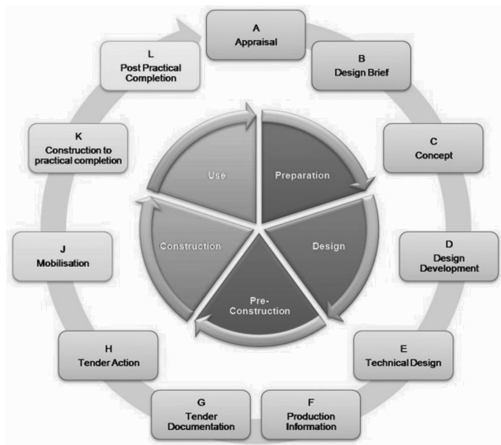
Figure 1. RIBA Outline Plan of Work.

PRACTICE ORIENTED BIM FRAMEWORK AND WORKFLOWS

This research conducted in collaboration between industry and academia aimed at developing a practice-oriented BIM framework and workflows that consider BIM processes and methodologies alongside technologies throughout the lifecycle of a building project (e.g. design, construction and operation). The core concepts underpinning the methodology adopted consists of the following:

- Exploit the domains of knowledge and concepts identified in BIM theoretical frameworks as a point of departure;
- Use those concepts to elicit knowledge and understanding from stakeholders, involved in building projects, to create a BIM framework and workflow;
- Consider BIM workstreams of all project lifecycle phases and the corresponding BIM technologies and their interoperability;
- Align the BIM framework and workflow with the standard design processes adopted in the UK (i.e. RIBA Outline Plan of Work).

For each of the element of the above methodology, the project utilized a specific research method. The literature review was utilized to review previous BIM frameworks and workflows. Interviews were utilized to elicit knowledge from industry professional. Case studies aimed to test the effectiveness of the BIM framework and workflow developed.

From the literature review, it was concluded that there is still a gap in terms of BIM frameworks and workflows that can be implemented by organizations in the building industry at project level. However, there are some theoretical frameworks (i.e. Bilal Succar, 2009) that defines in a concise manner the concepts and domains of knowledge required for BIM implementation in the industry and could be used as a theoretical point of departure.

In the interviews, the BIM fields (i.e. technology, process, policy) identified in an established theoretical framework – presented by Bilal Succar (2008), were used to elicit knowledge and understanding from major stakeholders in the building industry.

The knowledge and understanding obtained from these stakeholders was used to define the system specifications for the BIM framework and workflows. A total of 16 semi-structured interviews with individuals from major consultant, client and contractor organizations were conducted. Interviewed members represented nine large consultants and architects, two major city councils representing the client, and four large and multinational contractors. Members were included in the sample only if they are considered "experts" in both BIM concepts and business processes.

Table 1. Common Themes of the System Specification.

Information	Technological Framework	Contractual Agreements
BIM protocols	Interoperability	Responsibilities
Lifecycle Integration	Standardised Content	BIM Champions
Project Execution plans	BIM Services	Best Practice Guidance
Implementation in projects	Sustainability	BIM Awareness

Table 2. Elements Related to each Theme.

Themes	Elements
Information	- Communication - Structuring - System - Interoperability - Security
Lifecycle Integration	- RIBA stages - BIM workstreams
Project Execution Plan	- Roles and responsibilities - Model use workflows - Deliverables - Levels of details - Software selection - Model access - Technology assessment
Technological Framework	- Design - Analyse - Manage - Review - Case studies - Infrastructure - Software specific manuals
BIM Protocols	- Templates - Style Guide - Libraries - Best practices
BIM Champions/Coordinators	- Role and Responsibilities
Contractual Agreement	- Integrated Project Delivery - Agreements
Implementation in projects	- BIM workshop - BIM execution plans - BIM process approval

The initial stage of requirement capture consisted in posing broad trigger questions about the requirements of BIM frameworks and workflows (i.e. processes, technologies, policies). The qualitative analysis of the answers given involved the organization, classification and coding of answers, and have led to the identification of common themes (Table 1) mentioned by more than 90% of the interviewees. These themes were then discussed with an additional round of interviews to identify their elements and understand the interrelations between the elements. Table 2 shows the elements which are pertinent to each of the themes. The elements of these themes were further developed into a full set of specifications which also involved defining the interactions between elements from the different themes. An example of the interactions between BIM technology, BIM workstreams, interoperability and the RIBA Outline Plan of Work is mapped in figure 2. It was also recognized by interviewees that there is a need to develop a process that organizes and tailor the steps for the deployment of the BIM framework and workflows at individual project level. This process was also mapped and agreed by all the stakeholders (figure 3). It is composed of four phases that aim to assess feasibility of the execution of the BIM framework and workflows and tailor it to specific project needs.
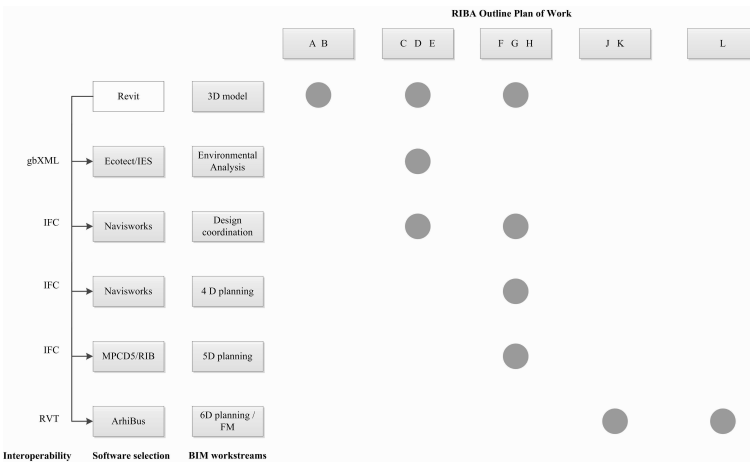
Figure 2. An example of interactions between elements of the BIM framework and their alignment with RIBA process.
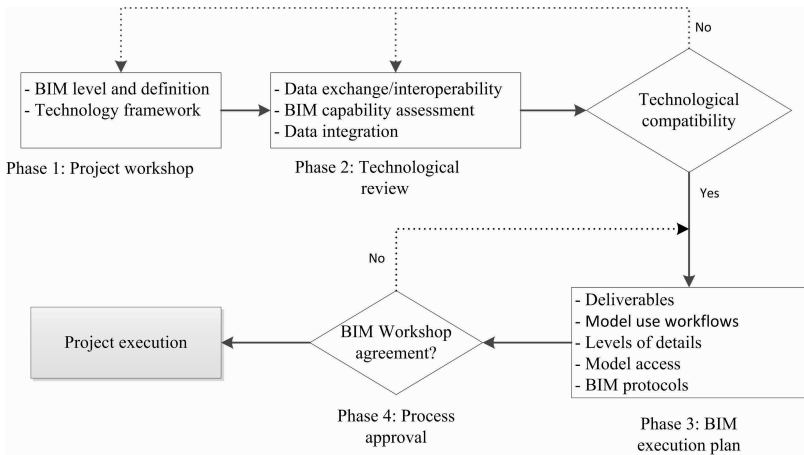
Figure 3. Process for the implementation of the BIM framework at project level

CASE STUDY

The developed BIM framework and workflows were tested in two international live design competitions called "Build London Live 2009" and "Build Qatar Live 2012" respectively. In these competitions, a number of teams composed of architects, engineering consultants (Mechanical, Electrical, Plumbing, Structural) and contractors compete in a 48-hour to deliver a project design which is released just before the competitions starts. The winner is judged by experts from the AEC industry, major software vendors and organisations such BuildingSMART on the basis of the following criteria: compliance to the brief; design impact and clarity; Multi-disciplinary BIM and use of Interoperability, and use of BIM for technical assessment. In Build London Live 2009, the context of the brief was based on a hypothetical manmade island, whereby teams were required to design a mixed use scheme; including hotel, office, retail and residential space. The BIM framework explained earlier was followed and the project started with a workshop where a technology review (figure 4) was established followed by an analysis of the model use workflow and technological compatibility analysis (figure 5). Once the team agreed on the BIM feasibility assessment, the project execution started. The main deliverables were mono-disciplinary models; the integrated and coordinated multi-disciplinary model; an environmental assessment of design options, and a construction simulation model. Our team followed the presented BIM framework and workflows and started with the Architect defining the model coordinates. Then, the grid and floor plates were established and communicated to the structural and MEP engineers. The completion of the multi-disciplinary models (figure 6) was completed in a timely fashion with teams using various technologies and exchanging models via a BIM server. The architectural model was then adapted in such a way that areas and spaces are correctly configured to support downstream activities such as environmental analysis, design coordination and 4D planning (figure 6). Within the

competition time (i.e. 48 hours) all the BIM deliverables were produced by the team following the proposed BIM framework. The judges awarded our team the prize for the 'Best Multi-disciplinary BIM and use of Interoperability' in Build London Live in 2010 and the overall winner in a recent competition (Build Qatar Live 2012). Although the judgment is made only on the basis of general criteria assessed by expert judges, the two design competitions, where there is a high pressure to produce multidisciplinary designs in a very short time, showed that the proposed BIM framework and workflows help creating a shared vision about the implementation of BIM; facilitating communication and workflow, and increasing the efficiency of all stakeholders involved in the design process.
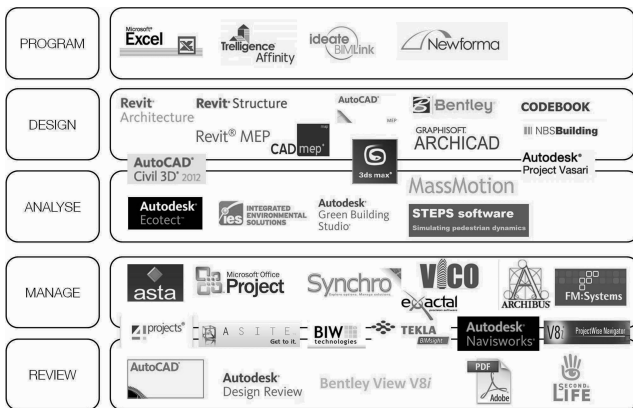
Figure 4. Technology framework
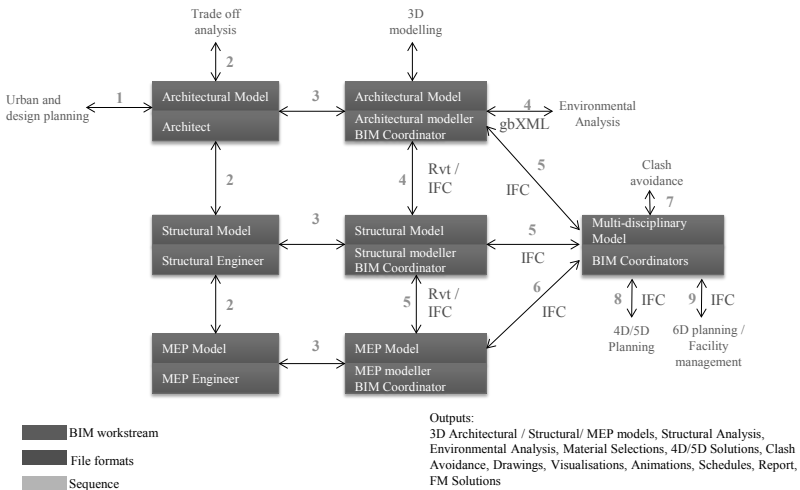
Figure 5. Model Use Workflow and Technological Compatibility
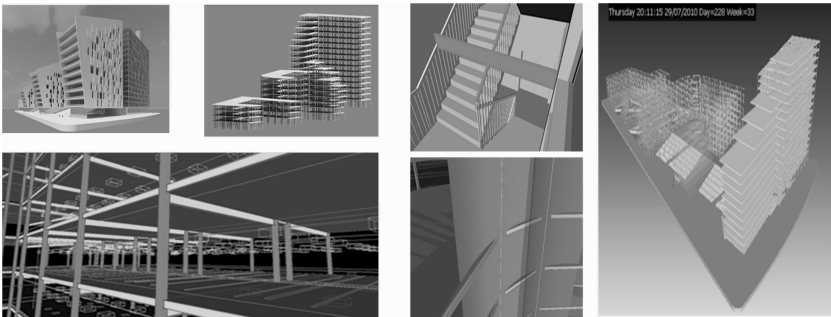
Figure 6. Architectural, Structural and Multi-disciplinary Models (left) and Design coordination and 4D planning model (right)

CONCLUSION

This paper presented a practice-oriented BIM framework and BIM workflows developed in a joint effort between industry and academia. The BIM framework consisted of multi-dimensional BIM processes that are utilized alongside BIM technologies throughout the lifecycle of a building project and are aligned with the standard design processes adopted in the industry. The requirements of the BIM framework were captured from two rounds of sixteen interviews with stakeholders representing all the trades involved in the design phase of buildings. The BIM framework consisted of a number of themes (i.e. information, lifecycle integration, project execution plan, technological framework, BIM protocols, BIM

champions/coordinators, and contractual agreement) with their detailed elements and interrelationships and a process for its implementation in projects. The developed BIM framework and workflows were tested in two time-constrained international design competitions and proved to be an effective tool in increasing the efficiency of the workflow of the design process between project stakeholders.

This BIM framework is currently implemented on many projects in the UK through consultancy delivered by a spin out company called 'BIM Academy' owned by Ryder Architecture who co-financed this research project.

Future studies will develop Key Performance Indicators (KPIs) to assess the role of the BIM framework and workflows in homogenizing the BIM maturity levels among the different trades involved at the design stage. This need resulted from observing some trades in the design team, who had less maturity and knowledge in BIM than other trades, increasing their understanding of BIM processes and technologies as the project proceeds.

REFERENCES

AGCA - The Associated General Contractor of America (2007) *The Contractors' Guide to BIM*, Publication No. 2926, edition one, p. 3.

BCA - Building and Construction Authority (2012) *Singapore BIM Guide*, http://www.corenet.gov.sg [last accessed November 2011]

BSI – Department of Business Skills and Innovation (2012) PAS 1192-2:2012 *Building Information Management – Information requirements for the capital delivery phase of construction projects*, http://www.cic.org.uk [last accessed Jan 2011]

EASTMAN, C. et al. (2011) *BIM Handbook: A Guide to Building Information Modelling for Owners, Managers, Designers, Engineers, and Contractors,* 2nd Edition, New Jersey: John Wiley & Sons, Inc.

New York City Department of Design & Construction (2012) *BIM Guidelines*, Long Island City, NY.

NIST - National Institute of Building Sciences (2007) *National Building Information Modeling Standard – part 1: Overview, Principles and Methodologies*, Washington, DC, U.S.

Penn State (2010) *BIM Project Execution Planning Guide and Templates – Version 2.0 BIM Project Execution Planning*, CIC Research Group, Department of Architectural Engineering, The Pennsylvania State University.

RIBA - Royal Institute of British Architects (2012) *BIM Overlay to the RIBA Outline*, RIBA Publishing, London.

Succar, B. (2009) *Building information modeling framework: a research and delivery foundation for industry stakeholders*, *Automation in Construction*, 18 (3), pp. 357-75.

US General Services Administration (2008*) 3D-4D Building Information Modelling, Series 04 - 4D Phasing*, Washington DC, U.S.

USACE - U.S. Army Corps of Engineers (2010) *USACE BIM Project Execution Plan*, Version 1.0, Washington DC, U.S.

As-Built Error Modeling for Effective 3D Laser Scanning on Construction Sites

Z. Shen[1], P. Tang[1], O. Kanaan[2], Y. K. Cho[3]

[1] Del E. Webb School of Construction, Arizona State University, 651 E. University Drive, Tempe, AZ 85287-0204; PH (480) 727-8105; FAX (480) 965-1769; email: zshen8@asu.edu, tangpingbo@asu.edu
[2] Department of Civil and Construction Engineering, Western Michigan University, 4601 Campus Drive, Kalamazoo, MI 49008-5316; email: kanaan.omar@gmail.com
[3]Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska-Lincoln, Peter Kiewit Institute (PKI) 104C, 1110 S. 67th St., Omaha, NE 68182; email: ycho2@unl.edu

ABSTRACT

Construction engineers increasingly use 3D laser scanners on job sites to capture detailed as-built geometries for construction monitoring, while their data collection time and quality can vary significantly. Various data collection parameters influence the quality of data and time for data collection. These parameters include internal (e.g., the laser signal strength, spatial resolution of the used scanner) and external parameters (e.g., color, object dimension, distance, and lighting condition). Correlations among these parameters, data collection time, and as-built modeling errors, contain principles of efficient and effective laser scan planning. Limited quantitative studies of these correlations result in ad-hoc laser scan planning, missing geometric information, and unnecessarily long data collection time. This paper presents a series of controlled experiments and seven statistical methods for quantitatively modeling the correlations between 3D as-built modeling errors and various parameters mentioned above. The major findings include: 1) internal correlations among data collection parameters need to be resolved for precise as-built error prediction; 2) all tested statistical methods are based on linear regression, while non-Gaussian distribution of the as-built modeling errors and nonlinear correlations in the experimental results distort the parametric error models. Future research will explore nonlinear or non-parametric methods for obtaining more reliable error models for proactively guiding engineers to conduct laser scanning in the field.

INTRODUCTION

Recent technological advances give engineers the ability to collect large amounts of detailed as-built geometric information/data on physical objects in the field. One such technology is laser scanning. A laser scanner can collect millions of individual 3D points in several minutes, forming dense 3D data sets known as point

clouds (Cheok et al. 2001). While laser scanners are capable of collecting dense point clouds with mm-level accuracy under lab conditions, construction sites contain various impediments that may compromise the quality of point clouds. Previous studies explored how various factors (e.g., angle-of-incidence, reflectivity) influence the positioning errors of individual 3D measurements or predefined targets (Cheok et al. 2011). On the other hand, limited understanding has been achieved about the relative importance of various factors influencing the qualities of surface models and geometric primitives (e.g., cylinders) extracted from point clouds, named "as-built modeling errors" in this research. As-built modeling errors are generally influenced by object properties (Boehler et al, 2003), time and space constraints for scanning (Boehler et al, 2003), and data density (Kiziltas 2008).

Object properties can significantly influence the as-built modeling error. Laser scanners detect the location of an object based on the laser signal reflected back from it. The interaction between the laser wave and the object surface, therefore, influences the quality of the collected data points. The reflective ability (albedo) of object surfaces influences the strength of the returning signal where, typically, white surfaces return stronger signals than black ones (Boehler and Marbs 2007; Tang et al. 2009). Moreover, focusing on minimizing the interruptions of construction activities under time and space constraints (Akinci et al. 2006; Latimer et al. 2004), limited scanning density can also compromise the data quality. In relation to time and space constraints on job sites, the data density of point clouds vary with the technical properties of the used scanner, i.e. the data collection rate and angular resolution, the scanning geometry, i.e. the incidence angle and the distance of the object from the scanner (Soudarissanane et al. 2007). Data densities have deterministic relationship with the accuracies of the geometric primitives extracted from point clouds (e.g., planes extracted through least-squares fitting), also known as "as-built modeling errors" (Vosselman et al., 2001).

Studies discussed above identified and characterized certain data collection variables as factors influencing the data quality of point clouds. Data collection variables of scanner parameters, environmental and geometric conditions are studied separately or in a combined manner (Soudarissanane et al. 2007; Tang et al. 2007). However, quantitative studies on comparing and prioritizing the impacts of a variety of physical and environmental variables on 3D scanning data quality are still limited. Knowledge about how to predict the as-built modeling errors given the scanning parameters and environmental conditions are also limited. Acquiring such knowledge will ensure a more cost effective 3D data collection process and reduce interferences of data collection activities with the construction process in the field.

The following sections begin with a short introduction of seven statistical analysis methods used in this paper (section "Methodology") to prioritize data collection variables of laser scanning. The detailed setup of controlled experiments and studied data collection variables are then presented (section "Field Tests Setting and Data Processing"). The sample data collected and processed are then used as inputs to those seven statistical methods for prioritizing the impacts of various data collection variables. Based on those prioritized variables, two error prediction models based on multi-linear regression are constructed, analyzed and compared. The one constructed by Rank Regression is recommended. Then, we discussed the effects of

various nonlinear factors not considered in the error prediction model based on multi-linear regression analysis, and described how the non-Gaussian distributions of some data collection variables cause challenges of error modeling based on linear models (section "Analysis"). The last section summarizes the research findings and identifies future research directions (section "Conclusions and Future Research").

METHODOLOGY

The authors investigate and compare results generated by seven statistical methods to identify and prioritize data collection variables in terms of their impacts on the data quality of point clouds. The seven statistical methods are: 1) Pearson's Correlation (Dalgaard 2008); 2) Ordinary Least Regression (OLR); 3) Huber M-Regression; 5) Rank Regression; 4) Least-Absolute-Deviation (LAD) Regression (Birkes and Dodge 1993); 6) Akaike Information Criterion (AIC) based statistical model selection; and 7) Bayesian Information Criterion (BIC) based statistical model selection (Kuha 2004).

The underlying theory of all these statistical methods is "linear regression." As shown in equation (1), they all assume that the response variable y_i is the linear combination of explanatory variables $\mathbf{x}_i$.

$$y_i = \boldsymbol{\beta}' \mathbf{x}_i + \varepsilon_i \tag{1}$$

Where y_i is a response variable; $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, ..., \beta_m]'$ is the vector of regression coefficients with m explanatory variables; $\mathbf{x}_i = [\mathbf{1}, x_i^0, x_i^2, ..., x_i^m]'$ is the vector of explanatory items; ε_i is unexplainable error or "residual", and $\mathbf{i} = \mathbf{1}, ..., n$ enumerates the n observations of y. Details of these methods are in the references.

EXPERIMENT SETUP AND DATA PROCESSING

For characterizing the influence of various data collection variables on the scanning data quality, the authors designed a series of controlled experiments using nine cylinders of three different radii (4.4, 6.3, and 8.8 cm), each of which are of three different colors (white, grey, and black), as shown in Figure 1a. The shape and color of the objects was selected to capture a wide range of surface reflectance (white surfaces reflect higher percentages of laser energy back), and understand how surfaces of different curvature influence the as-built modeling accuracy. The 3-D point clouds were collected by a Time-of-Flight (TOF) scanner.

Scanning data were collected at distances of 5, 10, 15, 20, 30, 40, 50, 70, 90, 120 and 150 meters. These are the data collection distances used in many construction sites. At each distance, we conducted low (0.2 m sampling spacing at 100 m), medium (0.1 m at 100 m), high (5cm at 100 m), and ultra-high resolution (2 cm at 100 m) scans, except at 15m and 30m: only low and medium resolution scans are collected at these two locations. The same set of scans was repeated during the daytime and nighttime to analyze the impacts of lighting. Hourly weather data, containing temperature, humidity, and dew point, was also collected from the National Oceanic and Atmospheric Administration (NOAA) for the time period of the experiment and was interpolated to match the exact times of data collections.

<center>(a)   (b)</center>
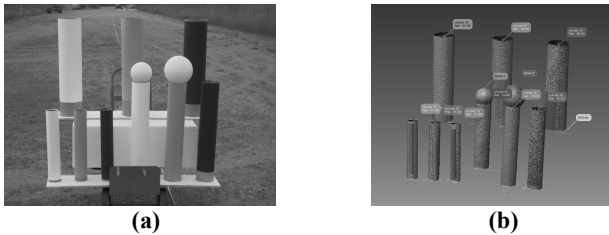
Figure 1. (a) Laser Scanning Data Collection Experimental Setup; (b) Screenshot of Sample Radius Extraction Using 3D Reverse Engineering Environment

Once all 80 scans were collected, a 3D reverse engineering environment was used to extract the radii of the cylinders. An example of the extracted cylinders was shown in Figure 1b. This radius extraction process was repeated three times to understand how uncertainties of manually selecting the points to be fitted influence reliability of the radius extraction results. An examination of the three sessions of manual radius extraction on all daytime scans is summarized in Table 1. These *average variations* were calculated for all cylinders at a given distance and resolution by finding the maximum difference between any of the three extracted radii for any given cylinder. As can be seen from this table, the average variations of the three manually extracted radii of all cylinders are all less than 0.1 cm, and have an average of 0.067 cm, indicating that the impacts of manual data selection is relatively small.

Table 1. Average Radius Extraction Variations by Size, Color, and Resolution

	Size			Color			Resolution			
	Large	Small	Med	White	Gray	Black	Low	Med	High	U.High
AV (mm)	0.050	0.062	0.090	0.043	0.073	0.089	0.070	0.072	0.063	0.067

* AV = Average Variation, Med = Medium, U. High = Ultra High

Quantitative statistical analyses require converting non-numeric values into numeric values. Items such as day, night, low resolution, medium resolution etc. were substituted with numbers. We also substituted error values in cases where the radius was un-extractable. As the scanning distances increased the data density decreased, leading to un-extractable radii beginning with black and small cylinders. Figure 2 shows the loss of extractable radii as a function of distance. In this research, when a radius was not extractable, the maximum as-built modeling error measured at all distances for that given cylinder was used. That is, if cylinder 5 (C5) at 50 m was un-extractable, the maximum extracted error for C5 among all other cases was used.

ANALYSIS

We conducted three types of analyses on the radius errors and data collection variables on all as-built radius error cases (examples shown in table 2): 1): Variable Ranking: prioritize data collection variables in terms of their impacts on the radius errors; 2) As-Built Error Modeling: create and evaluate error prediction models that predict the radius error given the values of data collection variables; and 3) Nonlinear

Factors Analysis: analyzing the performance of as-built error prediction models and evaluating the impacts of nonlinear factors on the performance of these linear-regression based error prediction models.
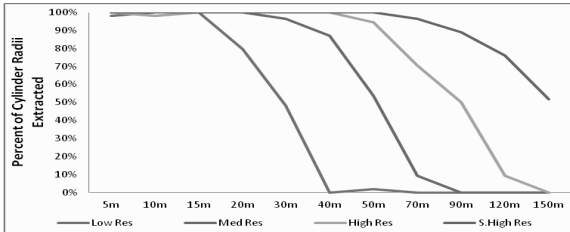
Figure 2. Cylinder Radius Extraction Rates at Various Distances

Table 2. Sample of As-Built Radius Error Cases Used for Statistical Analysis

O	D	T	R	C	MR	TR	T	DP	RH	E
C1	5	1	0	0	8.91	8.85	80	61	52	1.27
C9	70	0	3	2	6.23	6.32	73	58	59	2.39

* O = Object (C1: Cylinder 1, C9: Cylinder 9), D = Distance (m), T = Temperature (Fahrenheit), R = Resolution (Low = 0, Medium = 1, High = 2, Super High =3), C = Color (White = 0, Gray = 1, Black = 2), MR = Measured Radius (cm), TR = True Radius (cm), T = Temperature (F°), DP = Dew Point, RH = Relative Humidity (%), E = Error (cm)

Analysis I: Data Collection Variables Ranking

Explanatory variables (data collection variables) with smaller p-values (Jay L, 2011) are believed to have higher impacts on the response variable (radius error). Table 3 shows the data collection variables ranking results based on their p-values.

Table 3. Summary of Rankings Obtained from Statistical Analysis

Rank	Pearson	OLR	Huber	LAD	RR	AIC	BIC
1	*I*	*D*	*R*	*D*	*R*	*I*	*I*
2	*D*	*R*	*D*	*R*	*D*	*R*	*R*
3	*R*	*C*	*I*	*Rd*	*I*	*D*	*D*
4	*C*	*I*	*Rd*	*I*	*Rd*	*C*	*C*
5	RH	DP	*C*	RH	*C*	RH	DP
6	T	T	DP	DP	DP	DP	RH
7	Rd	Rd	T	T	T	Rd	T
8	DP	RH	RH	C	RH	TD	TD
9	TD	TD	TD	TD	TD	T	Rd

* I = intensity, D = Distance, R= Resolution, C = Color, RH = Relative Humidity (%), T = Temperature (Fahrenheit), Rd = radius, DP = Dew Point, TD = Time of Day (1 for daytime, 0 for nighttime), Pearson = Pearson's Correlation, Huber = Huber M-Regression, LAD = LAD Regression, RR = Rank Regression

Previous research has characterized how the distance between the scanner and the targeted object influence scan quality (Soudarissanane et al., 2007; Tang et al., 2009). The decrease in data quality with an increase in distance mainly due to the longer traveling distance of signal, which increases the beam width while decreases the signal strength (Lichti and Jamtsho, 2006). The presence of Resolution in the top

influential factors also agrees with the previous research. Averaging the ranking results obtained from our analysis and previous research results, we prioritized top five variables as: 1) Distance; 2) Resolution; 3) Color; 4) Intensity; and 5) Radius.

Analysis Ⅱ: As-Built Error Prediction Model

The collected cases of as-built modeling errors, 2052 in total, are separated into a training set with 1938 samples and a testing set with 114 samples for creating and evaluating as-built error prediction models. Given the training data set, methods of Rank Regression and OLR are used for estimating the linear coefficients of an error prediction model, which is a linear function depicting the mapping between the values of data collection variables and the radius error. The error prediction model estimated by Rank Regression and OLR are shown in equations (2) and (3) below.

$$ERR = 0.531 + 0.006D - 0.167R - 0.038Rd + 0.0517C - 0.00275I \tag{2}$$
$$EOLR = 0.3106 + 0.0066D - 0.1918R + 0.108C + 0.0545Rd - 0.0031I \tag{3}$$

Where, ERR represents the as-built modeling radius error predicted based on the Rank Regression and EOLR represents the radius error predicted based on OLR.

While most of the coefficients of the above parameters generally agree with each other in equations (2) and (3), the coefficient of the data collection variable "Radius" does not. According to (Xi et al. 2001), the larger the radius (Rd) is, the smaller as-built modeling error will be. In this sense, the coefficient of "Radius" should be negative. The positive coefficient of Radius in equation (3) is unexpected, and it is possible that the outliers in the training data set distorted the results acquired by the OLR algorithm. Further comparing the fitting effects of equations (2) and (3) on the testing samples, we found that the accumulative residual distribution of Rank Regression are mostly around 0 cm while their counterparts of OLR are slightly more widely scattered as shown in Figure 3a. It seems that the error prediction model based on Rank Regression is more consistent with previous research results and having less variation in the testing results.
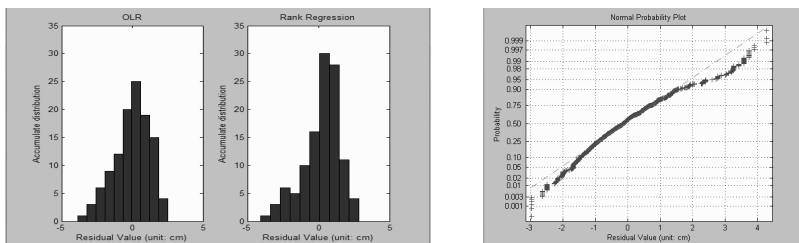
(a)   (b)

Figure 3. (a) Histograms of Residuals for OLR and Rank Regression Model; (b) Residual Normal Probability Plot of OLR

Analysis Ⅲ: Nonlinear Factor Analysis

All seven statistical methods studied herein are essentially based on linear regression methodology. While robust regression methods (e.g., Rank Regression) can tolerate outliers, they are still linear methods and cannot handle nonlinear relationships between the as-built modeling error and data collection variables. Error prediction model based on Regression Ranking does catch the first-order (linear) of a more complex error prediction model with higher orders (nonlinear). On the other hand, the relationship between as-built modeling error and the other five explanatory variables are, in essence, nonlinear, and some of these explanatory variables have internal correlations detailed in (Kanaan 2012). These internal correlations among data collection variables need to be resolved for developing more reliable as-built modeling error prediction models. Finally, figure 3b shows the normal probability plot of the residuals of OLR. This figure indicates that these residuals of error prediction do not completely follow the Gaussian distribution: residual values in the interval of [-1.5, 1.5] (unit: cm) seem to follow the Gaussian distribution, but the further from this interval, the residuals tend to less unlikely follow the Gaussian distribution. These results indicate that further studies are needed in understanding the distributions of as-built modeling errors in cylinders' radii.

CONCLUSIONS AND FUTURE RESEARCH

Various data collection variables influence the data quality collected by 3D laser scanner. Using statistical methods based on linear regression, this study quantitatively investigate the contributions of nine data collection variables on the as-built modeling error in radii of cylinders extracted from 3D laser scanned point clouds. The results indicate that the following five variables contribute the most to the radius errors: 1) Resolution; 2) Distance; 3) Color; 4) Radius; 5) Intensity. Using these five parameters as explanatory variables, we established two radius error prediction models based on Rank Regression algorithm and Ordinary Linear Regression (OLR). The results indicate that the Rank Regression algorithm is able to automatically eliminate the impacts of outliers in the experimental results, and produce error prediction results with less variation.

This research also analyzed the impacts on nonlinear factors on the reliability of the error modeling efforts presented in this paper. While all statistical methods used in this study are based on linear regression, the nonlinear factor analysis shows that strong nonlinear relations exist among data collection variables. Such nonlinear correlations among data collection variables need to be further investigated for more reliable ranking of the impacts of data collection variables. In addition, the normal probability plot of the error prediction residuals also indicate that the fundamental assumption of linear regression, "errors follow the Gaussian distribution," is not completely satisfied in the experimental results. Further nonlinear factor analysis is necessary for establishing more reliable nonlinear as-built error prediction models.

Future research will explore nonlinear statistical methods or non-parametric methods for obtaining more reliable as-built error models for proactively guiding engineers to conduct field data collections using the tested laser scanner. The authors also plan to conduct experiments using additional laser scanners to validate the generality of the research findings presented in this paper.

REFERENCES

Akinci, B., Boukamp, F., Gordon, C., Huber, D., Lyons, C., and Park, K. (2006). "A formalism for utilization of sensor systems and integrated project models for active construction quality control." *Automation in Construction*, 15(2), 124–138.

Birkes, D., and Dodge, Y. (1993). *Alternative Methods of Regression*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Boehler, W., and Marbs, A. (2007). "Investigating Laser Scanner Accuracy."

Cheok, G. S., Stone, W. C., Lipman, R. R., Witzgall, C., Bernal, J., Topical, I., Park, L., and Washington, S. (2001). "Laser scanning for construction metrology." (082).

Cheok, G. S., Saidi, K. S., Franaszek, M., Filliben, J. J., and Scott, N. (2011). *Characterization of the Range Performance of a 3D Imaging System*. Gaithersburg, MD.

Dalgaard, P. (2008). "Introductory Statistics with R." Springer New York, New York, NY.

Jay L, D. (2011). *Probability & Statistics for Engineering and the Sciences*. 328–342.

Kanaan, O. (2012). "Regression-Based Prioritization and Data Modeling for Customized Civil Engineering Data Collection." Western Michigan University.

Kiziltas, S., Burcu, A., Ergen, E., and Pingbo, T. (2008). "Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management." J. Inf. Technol. Constr., Special Issue on Sens. in Constr. Infrastruct. Man, ITcon, 13, 134–154.

Kuha, J. (2004). "AIC and BIC: Comparisons of Assumptions and Performance." *Sociological Methods & Research*, 33(2), 188–229.

Latimer, E., Dewitt, L., Saxena, R., Lyons, C., Smith, L. M.-, and Thayer, S. (2004). "with applications to construction environments." 4454–4460.

Soudarissanane, S., Ree, J. Van, Bucksch, A., and Lindenbergh, R. (2007). "Error budget of terrestrial laser scanning □: influence of the incide scan quality." 1–8.

Tang, P., Akinci, B., and Huber, D. (2009). "Quantification of edge loss of laser scanned data at spatial discontinuities." *Automation in Construction*, 18(8), 1070–1083.

Tang, P., Huber, D., and Akinci, B. (2007). "A Comparative Analysis of Depth-Discontinuity and Mixed-Pixel Detection Algorithms." *Sixth International Conference on 3-D Digital Imaging and Modeling (3DIM 2007)*, IEEE, Montreal, Canada, 29–38.

Vosselman, G., and Dijkman, S. (2001). "International Archives of Photogrammetry and Remote Sensing , Volume XXXIV-3 / W4 Annapolis , MD , 22-24 Oct . 2001 International Archives of Photogrammetry and Remote Sensing , Volume XXXIV-3 / W4 Annapolis , MD , 22-24 Oct . 2001." XXXIV, 22–24.

Xi, F., Liu, Y., and Feng, H.-Y. (2001). "Error Compensation for Three-Dimensional Line Laser Scanning Data." *The International Journal of Advanced Manufacturing Technology*, 18(3), 211–216.

Spatio-Temporal Progress Estimation for Highway Construction

Nazila Roofigari Esfahan[1], Antonio Páez[2], and Saiedeh Razavi[3]

[1]PhD student, Dep. Of Civil Engineering, McMaster University, Hamilton, Canada; email: roofing@mcmaster.ca

[2] Associate Professor, School of Geography and Earth Sciences, McMaster University, Hamilton, Canada; email: paezha@mcmaster.ca

[3] A.M.ASCE., Assistant Professor, Dep. Of Civil Engineering, McMaster University, Hamilton, Canada; email: razavi@mcmaster.ca

ABSTRACT

Highway construction involve repetitive, sequential operations that make them vulnerable to time and cost overruns. In order to increase their chance of being completed on time and within budget, these projects need efficient resource scheduling methods. This study presents a new scheduling system for highway construction projects. Location data generated from GPS technology can be used to track movement patterns of resources as well as progress of the project. Spatio-temporal constraints are used to generate dynamic project schedules and to visualize actual and potential project schedules and progress. By comparing observed versus planned progress, the system can anticipate delays. This information, in turn, can allow managers to implement adjustments in a flexible way to achieve better efficiency than post-hoc adjustments possible with traditional approaches.

INTRODUCTION

Highway projects are heavy construction works with highly equipment-intensive processes spread over miles. Like other linearly repetitive projects, highway works are characterized by sequential and repetitive operations and assignment of resources in a serial fashion, which implies scheduling, planning, and control challenges (Hegazy, 2005). Since the activities tend to be highly interrelated, delays at any stage can cause the overall project to go beyond schedule. In order to minimize potential issues, schedules developed for these projects should reflect the dynamic project environment, and be generated based on the most recent project performance data and also be updated in a timely manner (Song, et al., 2009). A number of methods have been proposed in the literature to develop linear schedules. Despite particular differences, all of these methods schedule project tasks by plotting the progress of repetitive activities against time, that is, their production rates. A shared limitation is that production rates are based on information available before execution, and there is limited flexibility to adapt to other possible production rates caused, for instance, by on-site conditions or reallocation of resources.

While production rates can be revised on a post-hoc basis, the emergence of location-aware technologies presents opportunities to improve site operation

management (Pradhananga and Teizen, 2013). Location aware technologies, such as Global Positioning Systems (GPS), Radio Frequency Identification (RFID), or Ultra Wideband (UWB) systems, can provide real time information about the location and movement patterns of resources. GPS in particular is a relatively inexpensive and widely accessible technology that can been used to track real-time locations of construction equipment such as earth-moving machines (Ackroyd, 1998; Navon & Shpatnitsky, 2005) and concrete-hauling trucks (Lu, et al., 2007). Adoption by construction firms allows for tracking of heavy construction equipment (e.g. BOT construction (www.botconstruction.ca), Keiwit (www.kiewit.com)). In highway projects, the movement of resources through locations is a key indicator of the project status, and therefore the information provided by location-aware technologies creates the opportunity to update project schedules based on actual movement and availability of resources. The use of location-aware technologies to oversee resource locations and job status at operation sites has been reviewed by different authors, from the perspective of progress evaluation, productivity analysis, and safety (Teizer, et al., 2008; Yang, et al., 2010; Yang, et al., 2011; Teizer & Vela, 2009). In recent studies, Gong and Caldas (Gong & Caldas, 2010) propound the use of location-aware technologies to extract information required for productivity analysis of construction processes, whereas Pradhananga and Teizer (2013) demonstrate the use of GPS for sampling the location of equipment in earth moving operations.

The contribution of the present study is to present a methodology for real-time tracking and updating of linear projects. Whereas the focus of Pradhananga and Teizer (2013) is on proximity analysis and zones of operation to prevent hazardous situations in earth moving operations, the focus of the approach presented here is on space-time management of resources for efficient linear scheduling. Gong and Caldas (2010), on the other hand, develop an automated algorithm to process video information. While they note that their technique can be used to extract productivity information, positional information and interactions between multiple resources is not considered. In contrast, the methodology developed in the present study directly deals with multiple resources.

Specifically method proposed in this paper is a flexible, and natural, evolution of linear scheduling techniques for the age of location-aware technologies. The method is based on: 1) the use of location-aware technologies for real-time tracking of resource; 2) the linear scheduling method (Harmelink & Rowings, 1998) as a resource-driven technique suitable for repetitive tasks; and 3) explicit consideration of spatio-temporal constraints to regulate and manage the necessary conditions for space-time convergence (or avoidance if desired for safety reasons) of moving resources.

PROPOSED METHOD

The method presented here integrates scheduling methods for highway projects with locational information of project resources. For this purpose, the method has two integrated phases, namely project scheduling before execution, followed by tracking and control of project progress using locational information. The method uses a modified form of the linear scheduling method (LSM) to generate the project baseline schedule. In

the modified form used here, space-time constraints are considered to not only schedule each activity, but also to consider alternative progress rates possible for the equipment executing them. The flowchart of the method is presented in Figure 1. The method then tracks and updates project schedules at certain pre-set time intervals taking into account real-time progress rates and actual relocation patterns of the resources.
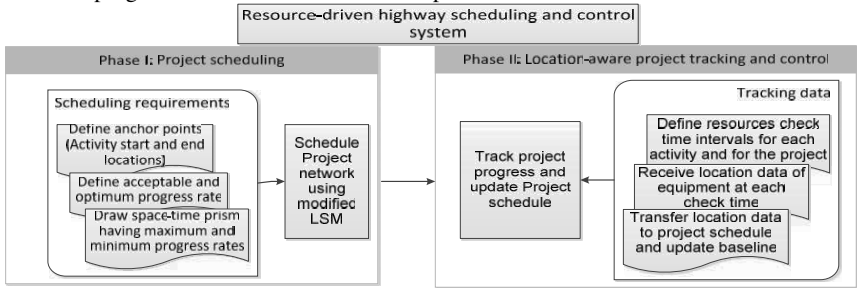
Figure 1: Proposed method

Space-time constraints are part of a conceptual framework widely used in the geographical analysis of moving entities (Miller, 2005). By considering potential trade-offs in the allocation of limited time among activities in space, space-time constraints can be used to assess the potential for space-time convergence of objects (Neutens et al., 2007). There are two key concepts. First, a space-time path traces the *actual* movement of an entity in space and time. A space-time prism gathers all *potential* space-time paths, given a time budget, and spatial and temporal anchors (locations and times where/when movement must begin/end). In terms of highway scheduling, space-time paths represent the progress of each activity over time; e.g. the kilometers excavated on a road. To generate the space-time prism for each activity, the start and end locations/times (anchors) of the activity are identified, and the maximum feasible production rates and floating times are used to generate the prism. Figure 4(a) illustrates an example of generated project schedule using modified LSM.
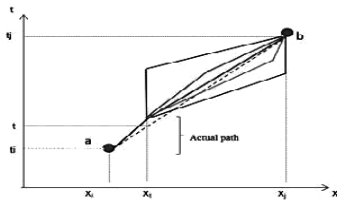
Figure 2

Figure 3

Figures 2 and 3: Updating Space-Time prism based on actual progress and project network for case example

After generating a baseline schedule, the method enters the tracking phase. Check time tracking intervals are selected by the project management team and are subjected to the duration of the project as well as managerial decisions on project tracking. As such, the intervals can approach real-time tracking as desired. In order to track and control the progress of the project, the method receives locational data gathered from GPS devices mounted on project equipment. These data are used to calculate actual progress rates during the relevant interval. The space time-prism for each activity is then updated substituting the actual progress rates with the space-time prism at that interval as shown in figure 2. Actual progress rates are compared with rates required to finish the activity at its pre-defined finish location and time. Depending on activity progress, alternative production rates can be suggested to finish the activity on time or to have the least possible delay for that activity. Next, the method is demonstrated using a numerical example.

APPLICATION AND NUMERICAL EXAMPLE

The method presented here was implemented in the MATLAB computing environment and was applied to a numerical example to demonstrate its capabilities. In the example, the management team is requested to enter general information about the project and its activities. The check time intervals in which the location data of equipment are to be updated also need to be provided. Based on these location data, the actual progress rates and the project status at each time interval are determined. By comparing these actual rates with those previously planned for each activity, the application suggests altering progress. The step by step application of the method is elaborated in detail in a numerical example.

The simple project network consists of 6 activities shown in Figure 3; including construction of a 4 kilometer altering access road and is planned to be finished in 15 days. Four activities (A, C, D and F) are assumed to be linear activities while the other two (B and E) were considered to be non-repetitive for this specific example. The check time intervals are considered being 3 days, meaning that the project schedule is to be updated based on the actual progress rates obtained. After entering project and activities information, the schedule of the project was generated, using linear scheduling method. In addition to the optimum planned progress rates for each activity, this schedule also includes each activity s space-time prisms (in this case, it is two-dimensional), taking into account its min and max progress rates. The overlap (congestion) areas between activities prisms were also identified.

The best case scenario, in which all activities are executed at their maximum progress rate along with its associated duration are generated from the input data (See Figure 5). This causes project to be finished 3 days ahead of time (i.e. in 12 days). This obviously shows the advantage of considering all possible progress rates while constructing the activity space time prisms. In generating the baseline schedule, Activities B and E are presented as horizontal lines, as they are considered not-repetitive. Figure 4(a) and 4(b) show the baseline schedule and best case scenarios for this project. As it is shown in Figure 4(a), activities C and D have the highest possibility of having resources space

conflicts, as their space-time prisms have the most overlap. Likewise, any deviation from planned finish time for activity D will result in delays in finishing of activity F as well.
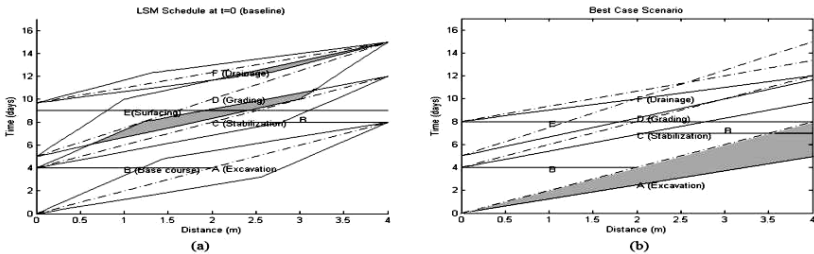
Figure 3: Example of: (a) generated project schedule using modified LSM and (b) best case scenario

After generating the baseline schedule for the project, project progress was simulated for each check time interval. For this purpose, at each check time, e.g. t=3 days, random numbers are generated to represent the location of the moving resources for the repetitive activities. These random numbers are generated form uniform distributions using MATLAB functions and are between minimum and maximum spatial boundaries of the activities under execution. As an example, activity A is the only activity executing at time t=3 days. At this time, based on activity A s optimum progress rate, this activity should progress 1.5 more kilometers. However, the progress could be anywhere between 0.9 Km at its slowest rate to 2.4 km at its fastest rate. To simulate the use of location-aware technologies, random numbers are drawn from the interval [0.9, 2.4] km and are used as observed location of the equipment. In real cases, the information on resource movements and their actual progress rate would be available from the use of location-aware technology. It should be noted that, in this approach it is assumed that no equipment idle time is observed between time intervals. In this context, being idle is equal to no change in the location of the equipment.

The actual progress rate for each interval is calculated considering the random number for the location. The activities space-time prisms are then substituted with the actual paths (Figure 6) at this interval and prisms are regenerated for the time between the current time and the project completion. Furthermore, the software suggest progress rate in which activity can finish at the date as planned. This way, a number of scenarios can be generated at each time interval, which will result in a better and more reliable project control system and easier decision making.

Two scenarios in addition to the best case scenario are presented here (Figures 5 and 6). The first scenario (Figure 5) shows that by taking the recommended corrective action, the project can still finish on-time regardless of different progress rates taken by equipment; i.e. when the progress rates taken are not necessarily similar to their optimum rate. In the second scenario (Figure 6), the project finished 2 days late on day 17th. As it can be seen in Figure 5, in scenario 1, activity A progressed slower than

expected in the first check time (Figure 5(b)). It should be noted that since the progress rate is calculated as the distance of the project executed per day, the slope of the lines in the generated prisms shows the inverse of the progress rate. This means that the more sloped a line is at a time interval, the less progress rate has been gained in that time interval. As a result, activity A should have a higher progress rate in order to be finished at its planned finish date (i.e. at day 8). As it is shown in that figure, because of taking corrective actions activity A could finish at its planned finish date although it did not progressed at its planned rate. In contrast, activity C failed to finish on time, because of not taking timely corrective actions, despite its faster starting rate than its planned rate. This obviously shows the importance of taking timely corrective actions in better control of highway projects.
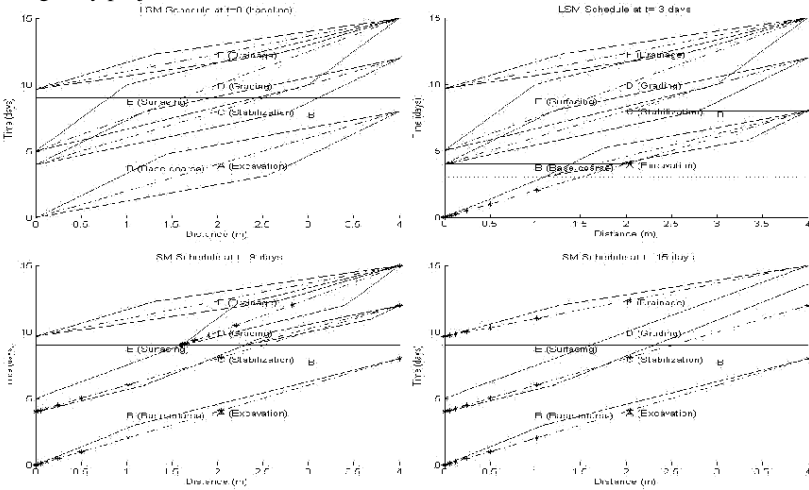
Figure 4: First Scenario, project schedule and its tracked progress at: (a) t=0 (baseline), (b) t=3days, (c) t=9 days and (d) t=15 days (final executed schedule)

This simple example demonstrates the ability of the proposed method. The method adds to the current scheduling methods for highway projects by a) considering all possible progress rates for each activity, and not only the optimum one; and b) constantly update project schedule, taking availability and movement of resources into account. This helps project managers to arrive in alternative plans to meet project deadline. By integrating location and movement data of resources with the project schedule, the proposed project control system is able to detect any delay, congestion, or activity idle time in near real time. As it has been shown through different scenarios, by knowing the location and relocation pattern of the resources of highway projects, and tracking such patterns throughout execution of the project, preventive actions can be performed before time and cost overruns happen. This way, delays and reductions in

construction progress are detected before they actually occur. Also through visual demonstration of the overlapped areas, the extent of any deviation from planned progress rates of each activity on its successors is more clearly shown, and can better assist the decision makers.
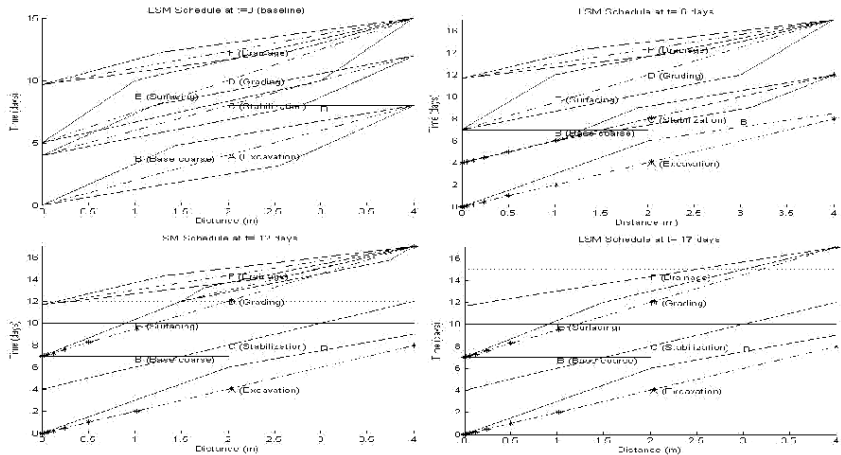
Figure 5: Second Scenario, project schedule and its tracked progress at: (a) t=0 (baseline), (b) t=6days, (c) t=12 days and (d) t=17 days

CONCLUSION

This study proposed a new progress tracking system for highway projects using location awareness in construction. This system expands the capabilities of current scheduling methods for highway projects. To do so, the use of space-time constraints along with GPS technology are proposed to assess the movement patterns of resources for each activity, and their potential for space-time convergence. Dynamic update of the location of resources and integration with a project schedule, more practical and flexible schedules can be obtained for more efficient project control.

The proposed method's capabilities were demonstrated using a simple numerical example. Two different scenarios as well as the best case scenario were considered. The best case scenario shows that understanding of all possible progress rates for activities facilitates decision-making to finish a project earlier than planned. The first scenario in which the project finished at pre-determined time budget, demonstrated the essence of the method in taking preventive action when delays are anticipated to happen as opposed to post-hoc remedies to past delays. The second scenario indicated the necessity of such control system to minimize delays to prevent cost overruns. The system can be beneficial for contractors and project managers of highway projects to track progress, to take proactive and efficient actions for project control.

REFERENCES

Ackroyd, N., 1998. Earthworks scheduling and monitoring using GPS. San Jose, California, Proceedings of Trimble Users Conference.

Fang, Z., Tu, W., Li, Q. & Li, Q., 2011. A multi-objective approach to scheduling joint participation with variable space and time preferences and opportunities. Journal of Transport Geography, 19(4), p. 623–634.

Gong, J. & Caldas, C., 2010. A computer vision based interpretation model for automated productivity analysis of construction operations. Journal of Computing in Civil Engineering, 24(3), pp. 252-263.

Hagerstrad, T., 1970. What about people in regional science?. Papers in Rigional Science, Volume 24, pp. 6-21.

Hegazy, T., 2005. Computerized system for efficient scheduling of highway construction. Transportation Research Record , Volume 1907, pp. 8-14.

Lu, M., Dai, F. & Chen, W., 2007. Real-time decision support for planning concrete plant operations enabled by integrating vehicle tracking technology,simulation, and optimization. Canadian Journal of of Civil Engineering, 34(8), p. 912–922.

Miller, H., 2005. A measurement theory for time geography. Geographical Analysis, Volume 37, pp. 17-45.

Navon, R. & Shpatnitsky, Y., 2005. A model for automated monitoring of road construction. Construction Management and Economics, 23(9), p. 941–951.

Sharma, H., McIntyre, C., Gao, Z. & Nguyen, T., 2009. Developing a Traffic Closure Integrated Linear Schedule for Highway Rehabilitation Projects. Journal of Construction Engineering and Management, 135(3), p. 146–155.

Song, L., Cooper, C. & Lee, C., 2009. Real-Time Simulation for Look-Ahead Scheduling of Heavy Construction Projects. Seattle, Washington, Construction Research Congress.

Teizer, J. & Vela, P., 2009. Personnel Tracking on Construction Sites using Video Cameras. Advanced Engineering Informatics, Special Issue, Elsevier, 23(4), pp. 452-462.

Teizer, J., Venugopal, M. & Walia, A., 2008. Ultra Wideband for Automated Real-Time Three-Dimentional Location Sensing for Workforce, Equipment, and Material Positioning and Tracking. Transportation Research Record: Journal of the Transportation Research Board, Volume 2081, pp. 56-64.

Pradhananga, N. and Teizer, J. 2013. Automatic Spatio-Temporal Analysis of Construction Equipment Operations using GPS Data, Automation in Construction, Elsevier, 29, 107-122.

Yang, J. et al., 2010. Tracking Multiple Workers on Construction Sites using Video Cameras. Special Issue of the Journal of Advanced Engineering Informatics, 24(4), pp. 428-434.

Yang, J. et al., 2011. A Performance Evaluation of Vision and Radio Frequency Tracking Methods for Interacting Workforce. Advanced Engineering Informatics, Elsevier, 25(4), pp. 736-747.

An Environment-Aware Sequence-Based Localization Algorithm for Supporting Building Emergency Response Operations

Nan Li [1], Burcin Becerik-Gerber [2], Bhaskar Krishnamachari [3], Lucio Soibelman [4]

[1, 2, 4] Sonny Astani Department of Civil and Environmental Engineering, University of Southern California; 3620 S Vermont Ave, KAP 210, Los Angeles, CA, 90089
[3] Ming Hsieh Department of Electrical Engineering, University of Southern California; 3740 McClintock Avenue, EEB 300, Los Angeles, CA, 90089

ABSTRACT

Building emergencies are big threats to the safety of building occupants and first responders. When emergencies occur, unfamiliar environments are difficult and dangerous for first responders to search and rescue, sometimes leading to secondary casualties. One way to reduce such hazards is to provide first responders with timely access to accurate location information. Despite its importance, access to the location information at emergency scenes is far from being automated and efficient. This paper identifies a set of requirements for indoor localization during emergency response operations through a nationwide survey, and proposes an environment-aware sequence-based localization algorithm that is free of signal path loss models or collection of prior data, and mitigates signal multipath effects. The algorithm enables efficient on-scene ad-hoc sensor network deployment and optimizes sensing space division by strategically selecting sensor node locations. Building information is integrated, in order to enable building-specific space divisions and to support context-based visualization of localization results. Proposed algorithm is evaluated through a building-size simulation. Room-level accuracy of up to 87.3% was reported, and up to 15.0% of deployment effort was reduced compared with using randomly selected sensor locations. The algorithm also showed good computational speed, with negligible time required for refreshing location estimation results in simulation.

INTRODUCTION

Building emergencies, including flooding, building collapses, terrorist attacks and especially structure fires, are big threats to the safety of building occupants and first responders. For example, public fire departments across the U.S. attended 484,500 fires in buildings in 2011, which caused 2,460 deaths and 15,635 injuries (Karter 2012). When emergencies occur, unfamiliar environments are difficult and dangerous for first responders to search and rescue, sometimes leading to secondary casualties. With the increasing number of complex buildings, and less live-fire training, first responders are twice as likely to die inside structures as they were 20 years ago, and the leading cause of these line-of-duty deaths is getting lost, being trapped or disoriented (Brouwer 2007). One way to reduce such hazards is to provide firefighters with timely access to accurate location information. It is also of critical importance for an incident commander to know the locations of deployed first responders in real time, so that decision-making process is faster and more informed.

When an emergency happens, first response teams are sent to carry out search and rescue operations. In most cases, searching for occupants is a manual process, which could be prohibited by fires, smoke or structural damage. Reducing the time spent on searching for occupants has great potential to reduce chances of fatalities and injuries.

LITERATURE REVIEW

Regardless of the high value of location information for building emergency response operations, current access to location information mainly relies on manual blind search by first responders. There are a few indoor localization solutions proposed in literature, but none has been widely adopted. Chandra-Sekaran *et al.* (2009a; 2009b) proposed a system to locate doctors and patients carrying radio nodes in outdoor/indoor emergencies. Monte Carlo and unscented Kalman filter techniques were used for location estimation. Accuracies between 5 to 10 m in simulation were reported. A system proposed by Duckworth *et al.* (2007) required no existing infrastructure or pre-characterization of the area of operation. The system relied on an ad-hoc network built on transmitters carried by both first responders in a building and vehicles outside the building. Cavanaugh *et al.* (2010) reported up to sub-meter accuracy with their system. The system required a considerable investment for on-site deployment of localization system-equipped vehicles. Rantakokko *et al.* (2011) proposed a system that integrated foot-mounted inertial sensors and Ultra Wide Band (UWB) sensors to support first responders. Field tests reported accuracy of 1 to 4 m. The system suffered from heading drifts. Akcan and Evrendilek (2012) proposed a system that utilized UWB technology. Directional localization was enabled in static networks. Reported accuracy through simulations was up to 6 m, depending on the node density. Another UWB-based system was proposed by Lo *et al.* (2008). It used a time difference of arrival (TDOA)-based algorithm for 3D location estimation, and reported accuracy of 1 to 2 m in field tests. The system required a significant deployment effort for a sensing network, and could not locate building occupants that had no access to mobile units. Kaya *et al.* (2007) used a backward ray-tracing algorithm to analyze angle of arrival (AOA), time of arrival (TOA) and signal power for locating first responders wearing beacons. Using multiple receivers, they were able to cover 80% of a building and achieve an accuracy of within 10 m.

There are also a few commercial solutions. Stemming from research sponsored by the Department of Homeland Security, SPIE's (Mapar 2010) solution, named "GLANSER", combined various technologies including global positioning system (GPS), IMU, UWB, Doppler radar, as well as a magnetometer, compass, pedometer, and altimeter inside a tiny wearable electronic unit. The algorithm was not disclosed, but an accuracy of 3 m was claimed in field tests. Exit Technologies (E2010) provided another solution that used handheld devices using low-frequency radios. A distressed first responder attempting reorientation or self-rescue could send out signals with a handheld device. Signals could then guide other first responders to the transmitting device. No details of the algorithm or accuracy were disclosed.

REQUIREMENT ANALYSIS FOR INDOOR LOCALIZATION

Most of the above solutions are highlighted by either their high accuracy or their independence from existing infrastructure. However, it remains unclear what

level of accuracy is sufficient to support emergency responses. Although a higher accuracy is desired, it may require a more sophisticated sensing network or additional prior data input. Independence from existing infrastructure is desired as it increases the robustness of a solution. However, robustness is also impacted by other factors, such as resistance to heat. These challenges are imposed by emergency scenes and require further examination. Prior research rarely discussed requirements other than accuracy and robustness. However, other requirements, such as computational speed, may be important to the success of emergency response operations.

To investigate indoor localization requirements for emergency response operations, an online survey was carried out. A list of eleven possible requirements was used in the survey (Table 1). The list was generated based on extensive discussions with first responders from the Los Angeles Fire Department (LAFD). A total of 1151 survey invitation emails were sent to first responders across the U.S. A total of 197 valid responses were received, which supported a ±6.8% confidence interval at a 95% confidence level. Participants had on average 25.7 years of experience, with all ranking levels from firefighters to fire chiefs.

Survey Results

Based on survey results, the requirements were organized in a descending order according to their importance in participants' point of view (Table 1).

Table 1: Importance of Indoor Localization Requirements

Rank	Requirement	% of Total Responses
1	Accuracy of location information	90.4%
2	Ease of deploying the solution on scene	83.8%
3	Resistance to heat, water and other physical damages	67.0%
4	Speed of calculating and presenting location information	66.0%
5	Size and weight of devices attached to first responders and occupants	58.9%
6	Purchase and maintenance costs	38.7%
7	Independence from building infrastructure (e.g. installed equipment) and building power supplies	22.8%
8	Independence from prior data collection (e.g. building layouts, and survey of radio features)	14.2%
9	Scalability of the solution to cover large numbers of people	14.2%
10	Ease of assembling the solution before dispatch	14.2%
11	Independence from on-scene data input (e.g. a few known locations inputted by first responders)	13.7%

Survey results showed that the most important requirements were: accuracy, ease of on-scene deployment, robustness (resistance to heat, water and other physical damages), computational speed (speed of calculating and presenting location information), and size and weight of devices. All of these five requirements were considered important by more than half of the total responders, which was remarkably higher than the percentage of all other requirements (13.7% to 38.7%). Accuracy was the foremost important, and participants indicated that room-level

accuracy was more desired than meter-level, floor-level or building-level accuracies. As measure of ease of on-scene deployment, participants reported that a maximum of 135 seconds was allowed to be spent on on-scene deployment. In terms of computational speed, an appropriate time reported by participants for data processing/location computation varied from 5 to 180 seconds, with an average of 40.34 seconds. Resistance to physical damages, and size and weight of devices are related to hardware, and therefore they are not in the scope of this paper.

EASBL ALGORITHM
Review of Sequence-Based Localization Algorithm

Sequence-Based Localization (SBL) is a range-free indoor localization algorithm (Yedavalli *et al.* 2005; Yedavalli and Krishnamachari 2008). It has a number of advantages that make it a desirable algorithm for satisfying the aforementioned indoor localization requirements. These advantages include capability of providing high accuracy, requiring low number of reference nodes, free of pre-data collection, and capability of mitigating multipath and fading effects.

At the heart of the SBL algorithm is the division of a 2D space into distinct regions. Consider a 2D space that consists of n reference nodes. For any two reference nodes, draw a perpendicular bisector to the line joining them. For n reference nodes, there are a total of $n(n-1)/2$ pairs and hence $n(n-1)/2$ perpendicular bisectors, dividing the space into a number of regions. For each region, an ordered sequence of reference nodes' ranks based on their distances to the region is defined as a location sequence of that region. Then, RSSI values of all reference nodes received by a target node are used to form the target node's location sequence. The centroid of a region whose location sequence is "nearest to" the target node location sequence is used as an estimated location of the target node. The nearness can be determined by e.g. Euclidean distance. The reference nodes and target nodes can be any type of radio frequency sensors that can communicate with each other.

Design of Environment-Aware Sequence-Based Localization Algorithm

Success of the SBL algorithm relies on the success of space division, which is essentially determined by the deployment of reference nodes. At emergency response scenes an ad-hoc sensor network must be quickly set up. There are a few challenges that must be addressed. Use of fewer reference nodes is crucial, as fast deployment is desired. In addition, SBL provides coordinate-level estimation. However, locations within the same region are not necessarily within the same room. This leads to a false room-level estimation. In other words, even when a coordinate-level accuracy is high, room-level accuracy may be low. Lastly, building elements such as walls impact accuracy and should be taken into consideration.

An Environment-Aware Sequence-Based Localization (EASBL) algorithm is proposed to address these challenges. EASBL measures the quality of space division with "breakaway area" p_{ba}. In SBL, the centroid of a region is used as an estimated location of a target node anywhere within that region. However, part of the region may be in a room different than the centroid, causing false room-level estimations. This part of the region is defined as a "breakaway area". A smaller p_{ba} within the sensing area indicates better space division and hence a higher room-level accuracy.

On-scene deployment effort is represented by the total number of reference nodes n, and by the difficulty in deploying each reference node. The difficulty in deploying reference node i is measured by penalty c_i, $1 \leq i \leq n$. There are two kinds of reference nodes: (1) hallway nodes (placed at hallway close to doors) are easy to deploy, and c_i is set to be 1; (2) room-center nodes (placed at centers of rooms) require more effort to deploy, and c_i is set to be 2. By using these candidate locations that do not need exact coordinates to be recorded or communicated, an incident commander can easily provide deployment commands to the first responses, and first responders can easily place the nodes and execute the commands.

Optimal ad-hoc sensing network deployment solution is one that minimizes the breakaway area and the penalty of all deployed nodes. From the computational point of view, this problem can be mathematically abstracted and expressed as: There are m candidate locations chosen based on building layout, and m reference nodes. Each candidate location $i (1 \leq i \leq m)$ can hold up to one node for deployment penalty c_i. Each node can be deployed at either one of the candidate locations or none of them (unused). For a given sensing area and given deployment of all nodes, a coverage penalty p_{ba} can be calculated based on the sensor locations and building layout. The objective is to minimize the total penalty (TP):

$$\min_{p_{ba}, \{c_i\}, \{k_{ij}\}} TP = e\, p_{ba} + \sum_{i=1}^{m} \sum_{j=1}^{m} c_i k_{ij}$$

$$s.t.\ c_i = \{ \frac{1, \text{if location } i \text{ is at hallway}}{2, \text{if location } i \text{ is at room center}}$$

$$k_{ij} \in \{ \frac{1, \text{if node } j \text{ is deployed at location } i}{0, \text{othersize}} \qquad (8)$$

$$0 \leq p_{ba} \leq 1$$

$$e > 0$$

where e is a coefficient balancing importance between the space division quality and the on-scene deployment effort, and k_{ij} is a binary variable that denotes whether a node j is deployed at candidate location i or not. Heuristics can be used for finding the optimal solution. As a widely used heuristic, a genetic algorithm is used in this paper. Other heuristics, such as simulated annealing and Tabu search, will be evaluated in future research.

Building information is used in three essential and critical ways in EASBL: (1) it is used to identify candidate locations for node deployment; (2) it lays the basis of calculating p_{ba} for a particular space division; (3) annotations such as room numbers can be used to facilitate quick recognition of a specific location for node deployment.

Simulation Setup and Scenarios

A C# script was written to implement EASBL. The script was compiled as a dynamic link library (DLL) file, and integrated into Autodesk Revit as an add-on. The add-on takes user input, extracts building geometries, performs space division optimization, and computes target locations. It then visualizes the estimated locations on floor plans. A simulation tool was programmed to simulate different localization

scenarios. It generates a number of targets in a sensing area and implements both a Random Placement SBL (RPSBL) and EASBL algorithms. It simulates the following signal propagation model (Rappaport 1996): $L(d) = L_0 + 10\gamma \log(d) + \sum_{p=1}^{P} WAF(p) + \varepsilon$, where $L(d)$ is path loss of signal strength (dB) in distance d (m), L_0 is reference signal strength loss in 1 m, γ is path loss exponent, WAF is wall attenuation factor, p is number of walls, and ε is a Gaussian term in log-normal fading. The values of L_0, γ and WAF used in simulation were 55.0 dB, 4.7 and 2.0 dB, respectively.

The fourth floor of the Ronald Tutor Hall (RTH) building on the University of Southern California campus was used as a simulation test bed. Two building fire scenarios with different scales were simulated. Both scenarios were designed based on suggestions from a number of first responders, and were verified by two battalion chiefs from the LAFD. In scenario 1 (Figure 1a), two single offices (red) were on fire. Occupants in both offices, all neighboring single offices, and offices and conference room that were across the hallway and had doors open to the hallway (orange) need to be evacuated. Due to the spreading smoke, visibility in the hallway outside the offices (cyan) was low, resulting in an increased risk to first responders. The sensing area is color-coded in Figure 1a with a size of 221 m^2. In scenario 2 (Figure 1b), a fire started in one lab and soon spread to a lab across the hallway (red). All labs on the east side of the floor were shut down for fire attack and search & rescue (orange). Visibility in the hallway was low due to smoke (cyan). The sensing area is color-coded in Figure 1b with a size of 729 m^2.

| (a) | (b) |
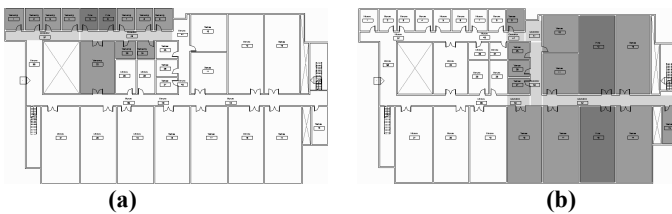
Figure 1: Simulation Scenarios

Simulation Results

In the simulation, a total of 50 targets (first responders and occupants) were randomly generated in the sensing area. Each scenario was simulated five times to offset the impact of randomness of target generation, and the simulation results were averaged. In addition, when running the genetic algorithm, every individual in the first generation was considered a random sample resulting from RPSBL, as their attributes were not impacted by crossover or mutation processes. All these first-generation individuals were averaged to get the results for RPSBL. Simulation results for both algorithms are presented in Table 2 for comparison.

The following four conclusions could be drawn based on these results. First, breakaway areas with EASBL were significantly lower than that with RPSBL in both scenarios, indicating a larger possibility of correct room-level estimation using the EASBL. Second, the total number of reference nodes to be deployed was generally comparable between two algorithms; however, a larger portion of reference nodes

had to be deployed at room centers with RPSBL, which pointed to a larger deployment effort. When the reference nodes were weighted with deployment penalty c_i, the total deployment effort with EASBL was 15.0% and 11.4% less than RPSBL in scenario 1 and scenario 2, respectively. Third, EASBL yielded both higher coordinate level accuracy and room level accuracy than RPSBL, with an overall improvement by 35.98% and 18.27%, respectively. Lastly, it was noticed that, after space division optimization was done, refreshing localization results took negligible amount of computational time with both algorithms, which was significantly less than 40.34 seconds, the maximum time allowed by survey participants.

Table 2: Indoor Localization Simulation Results

			RPSBL	EASBL
Scenario 1	Breakaway area (%)		24.8	7.7
	Sensor node deployment penalty	Room-center	9.2	7.3
		Hallway	2.3	3.0
	Average meter level accuracy (m)		2.43	1.52
	Average room level accuracy (%)		71.5	82.1
Scenario 2	Breakaway area (%)		19.3	9.2
	Sensor node deployment penalty	Room-center	10.3	8.4
		Hallway	3.1	4.2
	Average meter level accuracy (m)		2.46	1.81
	Average room level accuracy (%)		76.3	87.3

CONCLUSIONS

This paper identified a set of requirements for indoor localization at building emergency scenes. An EASBL algorithm was proposed to satisfy algorithm-related requirements. Results from a building-size simulation indicated that EASBL, while maintaining the advantages of SBL, was capable of addressing the challenges SBL had under emergency situations. The EASBL could serve to reduce on-scene deployment efforts and increase room-level accuracy, as desired by first responders when they carried out emergency response operations. In addition, since refreshing localization results could be done instantly in simulation, it suggested that the EASBL algorithm had a satisfying computational speed.

To further improve and evaluate the performance of EASBL, future research will be carried out to assess the impact of two parameters, including coefficient e and penalty c_i, on the optimization results and consequently on the localization accuracy and on-scene deployment effort. Parameter values of signal path loss models used in simulation also have impact on the evaluation results, hence deserving further examination. More importantly, the authors plan to perform real-world experiments, so that more comprehensive evaluation of the EASBL algorithm against all requirements including hardware-related ones can be carried out.

ACKNOWLEDGEMENT

recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Akcan, H., and Evrendilek, C. (2012). "GPS-Free Directional Localization Via Dual Wireless Radios." *Comput. Commun.,* 35(9), 1151-1163.

Brouwer, E. (2007). "Trainer's corner: What are your rules for calling mayday?" <http://www.firefightingincanada.com/content/view/1274/213/> (Jun 1, 2012).

Cavanaugh, A., Lowe, M., Cyganski, D., Duckworth, R. J. (2010). "WPI precision personnel location system: Rapid deployment antenna system and sensor fusion for 3D precision location." *Proc., Institute of Navigation - International Technical Meeting 2010, Jan 25-27, 2010,* 384-389.

Chandra-Sekaran, A., Weisser, P., Muller-Glaser, K., Kunze, C. (2009a). "A comparison of bayesian filter based approaches for patient localization during emergency response to crisis." *Proc., 2009 Third International Conference on Sensor Technologies and Applications (SENSORCOMM),* IEEE, 636-42.

Chandra-Sekaran, A., Stefansson, G., Kunze, C., Muller-Glaser, K., Weisser, P. (2009b). "A range-based monte carlo patient localization during emergency response to crisis." *Proc., 5th Advanced International Conference on Telecommunications, May 24-28, 2009,* IEEE Computer Society, 21-26.

Exit Technologies. (2010). <http://www.exit-technologies.com/draeger.php> (Apr 6, 2012).

Karter, M. J. (2012). *Fire Loss in the United States 2011,* National Fire Protection Association, Fire Analysis and Research Division, Quincy, MA.

Kaya, A. O., Greenstein, L., Chizhik, D., Valenzuela, R., Moayeri, N. (2007). "Emitter localization and visualization (ELVIS): A backward ray tracing algorithm for locating emitters." *Proc., 2007 41st Annual Conference on Information Sciences and Systems,* IEEE, 69-70.

Lo, A., Xia, L., Niemegeers, I., Bauge, T., Russell, M., Harmer, D. (2008). "EUROPCOM - an ultra-wideband (UWB)-based ad hoc network for emergency applications." *Proc., Proc. of VTC/Spring - 2008 IEEE 67th Vehicular Technology Conference,* IEEE, Singapore, 6-10.

Mapar, J. (2010). "Tracking emergency responders in challenging environments." <http://spie.org/x39740.xml?ArticleID=x39740> (April 6, 2012).

Rantakokko, J., Rydell, J., Stromback, P., Handel, P., Callmer, J., Tornqvist, D., Gustafsson, F., Jobs, M., Gruden, M. (2011). "Accurate and Reliable Soldier and First Responder Indoor Positioning: Multisensor Systems and Cooperative Localization." *IEEE Wireless Communications,* 18(2), 10-18.

Rappaport, T. S. (1996). "*Wireless Communications: Principles and Practice*", 1st edition Ed., Prentice Hall.

Yedavalli, K., Krishnamachari, B., Ravula, S., Srinivasan, B. (2005). "Ecolocation: A sequence based technique for RF-only localization in wireless sensor networks." *Proc. IPSN '05*, Los Angeles, CA, IEEE, 285-292.

Yedavalli, K., and Krishnamachari, B. (2008). "Sequence-based localization in wireless sensor networks." *IEEE Transactions on Mobile Computing,* 7(1), 81-94.

Application of Sensing Technology in the Prevention of Backing Accidents in Construction Work Zones

Sooyoung Choe[1], Fernanda Leite, M.ASCE[2], Dan Seedah[3], Carlos Caldas, M.ASCE[4]

[1]PhD Student, Construction Engineering and Project Management Program, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E Dean Keeton St. Stop C1752, Austin, TX 78712-1094; Ph (510) 529-8674; email: sooyoung.choe@utexas.edu
[2]Assistant Professor, Construction Engineering and Project Management Program, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E Dean Keeton St. Stop C1752, Austin, TX 78712-1094; Ph (512) 471-5957; Fax: (512) 471-3191; email: fernanda.leite@austin.utexas.edu
[3]Research Fellow, Center for Transportation Research, The University of Texas at Austin, 1616 Guadalupe Street, Suite 4.202, Austin, Texas, 78701; Ph (512) 232-3143; Fax: (512) 232-3070; email: dseedah@mail.utexas.edu
[4]Associate Professor, Construction Engineering and Project Management Program, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, 301 E Dean Keeton St. Stop C1752, Austin, TX 78712-1094; Ph (512) 471-6014; Fax: (512) 471-3191; email: caldas@mail.utexas.edu

ABSTRACT

Machinery and vehicle related fatalities represented 32% of construction fatalities in 2010; and backover fatalities were recorded as one of the main culprits due to limited visibility. In order to minimize backover accidents, sensor-based proximity warning technologies, such as video camera, radar, ultrasonic, and tag-based systems, have been developed, tested and in use in many new passenger vehicles. This paper addresses potential implementation of sensor-based proximity warning systems to minimize backing accidents by reviewing sensing technologies applicable for construction equipment and presenting a formalized framework of sensor-based proximity warning systems evaluation. The presented framework includes a generic sensor system selection approach and performance test bed protocol. The proposed framework is expected to help decision makers objectively evaluate sensor-based proximity warning systems in their selection to implementation in the construction work zones.

INTRODUCTION

In the 2010, 774 construction workers were killed in the United States. This number is more than 15% of overall work related fatalities in 2010, and fatal occupational injury rate is approximately 9.5 per 100,000 construction workers while overall 2010 fatal work injury rate is 3.6 per 100,000 full-time equivalent workers (BLS 2010). More than 32% of fatalities in the construction industry were caused by machinery and vehicles related accidents in 2010 (BLS 2010). The Occupational Safety and Health Administration (OSHA) indicated that about 360 workers were killed from backover accidents from 2005 to 2010 in the United States and are seeking solutions to prevent backing fatalities in construction and other industries (OSHA 2012).

To improve Texas traffic control practices, the Texas Department of Transportation (TxDOT) has been conducting a backover prevention research, entitled "Prevention of Backing Fatalities in Construction Work Zones." The main objectives of this research are to review current practice, identify commercially available technologies, and test commercially available systems for prevention of backing fatalities. The project identified four main strategies, which are site planning, signalers, training, and electronic devices to prevent backing fatalities in the construction industry. Among the four strategies, this paper focused on the use of electronic devices, specifically sensor-based proximity warning systems.

OSHA requires a backup alarm when a vehicle with a limited rear view is backing up (OSHA 2009). Although this requirement provides audible warning to workers on foot, there is a need to provide adequate warning to the operators who control the vehicles because of noise interference and the limitations of one-way communication. Caterpillar Inc. measured blind areas in several types of heavy construction vehicles to identify dangerous areas around vehicles. According to their research, blind areas can be reduced by applying sensing technologies (Hefner and Breen 2004). Many sensing technologies, such as radar, ultrasonic, and tag-based systems, have been developed and tested to provide hazard warnings to either operators or workers to prevent collisions within blind areas of heavy industrial equipment. Despite the potential benefits such technologies can bring to the construction work zone, very little data is available concerning the performance of these sensing technologies on actual construction equipment and work zones. Also, there is no 'ideal' sensing system satisfying various construction equipment types because required sensing system will vary according to the size type of vehicle, its blind areas, and governing reverse speeds.

The main objectives of this study are: (1) to review sensing technologies to prevent backing accidents in construction work zones; and (2) to present a formalized approach to evaluate and test sensor-based proximity warning systems.

REVIEW OF SENSING TECHNOLOGY

Several sensing technologies exist for collision detection and avoidance by alerting the driver with a visual and/or audible warning. In-vehicle sensing systems have emerged recently in the automotive market, but this paper will focus on

aftermarket sensing systems developed as backover collision avoidance purpose with the ability to warn drivers the presence of obstacles behind the vehicles because there were no commercially available construction vehicle models with installed sensing systems in the manufacturing phase. Potential sensing technologies explained this section include four technologies (e.g., video camera, ultrasonic, radar, and radio frequency identification) which are identified as backing safety solutions by OSHA (OSHA 2012), and global positioning system (GPS).

Video Camera

Video cameras are one of the most popular technologies to cover a vehicle's blind area. This system displays the rear view area of vehicles to operators via a cab-mounted monitor. The advantage of video camera is that this system is able to monitor the cause and exact location of obstacles, not just detecting the presence of obstacles. However, this system does not generate an alarm, so it is recommended as a supplemental method with other sensing technologies (Ruff 2000; Ruff 2001). Also, intensive glances are required to identify the nature of obstacles on the monitor and it is possible that a driver with a quick glance might fail to recognize a hazard behind a vehicle (Mazzae and Garrott 2007). Finally, their performance is questionable during limited visibility such as night time.

Ultrasonic

Ultrasonic-based sensing systems transmit sound waves (a typical frequency is 40 kHz) backward from the vehicle and detect echoes from an obstacle within a detection zone. The received echoes determine the distance to the obstacle and generate a visual and/or audible warning in the cab. This system is widely used with light vehicles to aid parking as well as low speed backing maneuvers and has relatively compact size and low price. However, ultrasonic systems have limited detection ranges, typically less than 3 meters (10 feet), and require multiple sensors to cover the width of large equipment (Ruff 2007).

Radar

Radar technology can be subdivided into two types, pulsed and doppler radars. Pulsed radar transmits pulsed signals within the predefined detection zones and detects an object by measuring the time of flight of a pulsed signal reflected from the object. Doppler radar uses the relative movement of an object to the vehicle (i.e., if both the vehicle and the object are stationary, the object is not detected). Both radar types are capable in effectively detecting people, other vehicles, and metal objects, but some materials, such as plastics and wood are not detected well by radar signals (ISO 2010; Ruff 2007). The primary concern in using radar-based sensing systems is frequent false and nuisance alarms which might eventually make the drivers start ignoring warnings of the systems, but this problem can be solved with use of supplemental methods such as video camera system (Ruff 2007). Also Ruff pointed out that radar and ultrasonic systems have sporadic detection (inconsistent detection)

areas at the boarder of detection areas (Ruff 2000; Ruff 2001). In addition, the detection performance of radar sensor system (also ultrasonic-based systems) is affected by sensor installation positions (Ruff 2007).

Radio Frequency Identification

Another type of proximity warning system (PWS) is a tag-based radio frequency identification (RFID). This system uses electronic personal tags equipped workers and a tag- reader detects tags within a detection zone. According to tag types, active and passive RFID systems exist. Active RFID systems allow two-way communication between an operator in the cab and a worker wearing a tag by generating warnings to both alert device in the cab and a tag. On the other hand, passive systems only generate an alarm to an operator when a worker with a tag presents within the detection zone. Despite very few false alarms, there is the potential of collisions amongst workers or objects that are not equipped with a tag (Ruff 2007). Since construction work zones are busy and dynamic, making all workers wear a tag is not realistic, especially open job sites. Also, tag-based RFID systems have relatively long detection ranges, but these systems are relatively expensive and hard to install than other systems (Ruff 2000; Ruff 2001). RFID systems have been widely studied for real-time monitoring and tracking in construction work zones (Maalek and Sadeghpour 2013; Marks and Teizer 2012; Naticchia et al. 2013).

Global Positioning System

Global positioning systems (GPS) use satellite signals and triangulation to detect objects. This system can cover wide areas but has accuracy problems with moving objects. Also, GPS does not provide accurate location estimates in areas near buildings and big trees that may interfere with signals from the satellites (Sacks et al. 2003). Due to accuracy issues, this system might not be appropriate for safety purposes in construction work zones which require a high degree of location detection accuracy.

SENSING SYSTEM SELECTION

Because of the various construction equipment characteristics (e.g., size and blind areas), there is no 'ideal' sensor system satisfying all construction equipment types. For instance, a pickup truck and a backhoe have a different vehicle size and blind areas which require unique sensor system selection considerations for each type of vehicle. Also, rapidly developing sensor markets creates the need for generic sensor selection criteria which can be applied when new sensor types are developed in the future. This research presents a formalized sensor selection approach, as illustrated in Figure 1.
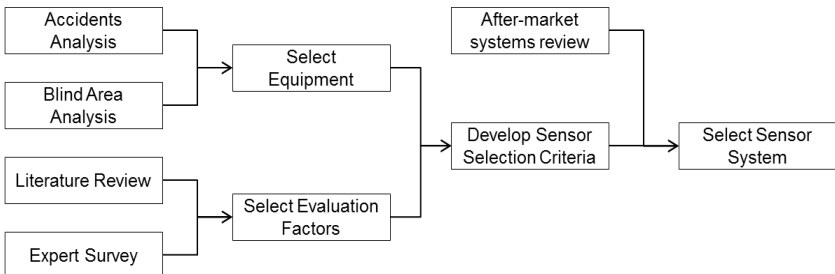
Figure 1. Logic approach for sensor system selection.

In order to generate generic sensor selection criteria for different construction equipment types, first, test bed equipment types are selected from the accidents analysis and examination of blind area of each equipment type. Also, from the previous research and survey conducted with experts in the safety or sensor related industry or academia, generic evaluation factors are determined. Based on possible sensor selection evaluation factors identified in literature, an expert survey was conducted to verify importance of each evaluation factor. With the sensor system's evaluation factors, sensor selection criteria for test bed equipment are developed. Lastly, proper sensor systems are selected by assessing after-market systems list and criteria developed for each test bed equipment type.

PERFORMANCE TEST

The performance test aims to present a test bed protocol that can apply to typical construction equipment. To access the performance of sensor systems selected in Section 3, the following experiments should be considered: Vehicle Measurement, Static Test, and Dynamic Test.

Vehicle Measurement is to consider proper sensor installation positions; at least test bed equipment types selected in Section 3 must be performed. Because each equipment type has different models with their own limited space for sensor installation, it is recommended to survey several types of models. In order to consider both technical and practical issues, it is also recommended that sensor installation positions reviews be conducted with system vendors and practitioners who are closely related to construction vehicles.

Static Test reflects the vehicle operation condition which is before the vehicle is backing (i.e., both a vehicle and an obstacle are stationary). In this scenario, a sensor system can generate a maximum size of detection performance regardless of sensor response time and vehicle operation conditions. The detection area measured in the best condition will serve as a benchmark for the dynamic test. Also, sensor detection performance at different installation locations measured from the vehicle measurement will be tested. To assess the performance of sensor systems in the static conditions, the following factors should be observed in the static test:

- Reliable Detection Area: This area indicates the detection coverage with more than 90% accuracy. This area will be used to determine overall detection capacity of the system.
- Sporadic Detection Area: This area indicates the detection coverage with less than 90% accuracy.
- Close Proximity Detection: This factor will evaluate sensor system's close proximity detection capability within 0.9 meter (3 feet) of vehicle and along the vehicle width.

Figure 2 illustrates an example of performance evaluation factors.
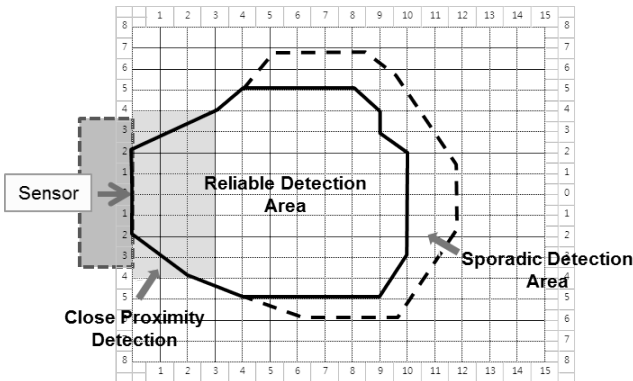
Figure 2. Example of performance evaluation factors in the static test.

During the dynamic test (see Figure 3), the dynamic detection ranges of sensor systems will be studied by placing sensor systems on moving equipment (i.e. a vehicle is backing and an obstacle is stationary). The impact of dynamic operation condition will be examined with possible environmental factors such as vehicle speed or sensor response time. Also, for the safety purpose, it is recommended using a mannequin as a test object, not human. To access the performance of sensor systems in the dynamic conditions, following performance evaluation factors will be checked:

- False alarm: alarms generated in the clear field because the sensor might detect the ground or parts of vehicle such as tires during the backing operation.
- Changed maximum detection range: changed detection range due to the dynamic conditions of the vehicle. This changed detection ranges should be compared to maximum detection ranges in the static condition to see differences.

Figure 3. Overview of the dynamic test.

CONCLUSIONS

In recent years, a number of advanced sensing technologies, such as video camera, ultrasonic, radar, RFID, and GPS have been developed as solutions in the prevention of backing accidents in construction as well as other work zones using vehicles frequently. However, there was no standardized testing and reporting protocols to evaluate sensor system performance specific to the uses of the construction industry. The paper reviewed several sensing technologies which might assist in equipment backing operation and sought to present a generic approach to selecting sensor-based proximity warning systems and evaluating their performance before actual use in work zones. This paper will have two possible main implications for researchers and practitioners in construction safety. Firstly, the research will present a formalized framework of sensor-based proximity warning systems evaluation from sensor system selection to performance test bed protocol specific to the construction industry. Secondly, contents of the research will provide users with a better understanding of the systems from installation to performance during static and dynamic construction environments. The next steps of the research will focus on the development of a test bed protocol which includes recommendations on the test area, objects, procedures, and data reporting for both static and dynamic work conditions.

ACKNOWLEDGEMENT

REFERENCES

BLS (2010). "Census of Fatal Occupational Injuries (CFOI) - Current and Revised Data." *Bureau of Labor Statistics*. <http://www.bls.gov/iif/oshcfoi1.htm> (Dec. 10, 2012).

Hefner, R. E., and Breen, P. J. (2004). "Construction Vehicle and Equipment Blind Area Diagrams." Center for Disease Control and Prevention, Contract #200-2002-00563.

ISO (2010). "Transport information and control systems – Maneuvering Aids for Low Speed Operation (MALSO) – Performance requirements and test procedures." *International Organization for Standardization*, ISO 17386.

Maalek, R., and Sadeghpour, F. (2013). "Accuracy assessment of Ultra-Wide Band technology in tracking static resources in indoor construction scenarios." *Automation in Construction*, 30, 170-183.

Marks, E., and Teizer, J. (2012). "Proximity Sensing and Warning Technology for Heavy Construction Equipment Operation." *Proc., Construction Research Congress 2012*, 981-990.

Mazzae, E. N., and Garrott, W. R. (2007). "Evaluation of the Performance of Available Backover Prevention Technologies for Light Vehicles." National Highway Traffic Safety Administration (NHTSA), 1-10.

Naticchia, B., Vaccarini, M., and Carbonari, A. (2013). "A monitoring system for real-time interference control on large construction sites." *Automation in Construction*, 29, 148-160.

OSHA (2009). "Regulations (Standards - 29 CFR)." *Occupational Safety & Health Administration*. <http://www.osha.gov/> (Jan. 9, 2013).

OSHA (2012). "Preventing Backovers." *Occupational Safety & Health Administration.* <http://www.osha.gov/doc/topics/backover/index.html> (Dec. 10, 2012)

OSHA (2012). "OSHA seeks comments on how to prevent worker injuries and deaths from reinforcing concrete activities and vehicle backovers." *Occupational Safety & Health Administration*. <http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=NEWS_RELEASES&p_id=22071> (Dec. 10, 2012).

Ruff, T. M. (2000). "Test Results of Collision Warning Systems on Off-Highway Dump Trucks." National Institute for Occupational Safety and Health (NIOSH), 1-50.

Ruff, T. M. (2001). "Test Results of Collision Warning Systems on Off-Highway Dump Trucks: Phase 2." National Institute for Occupational Safety and Health (NIOSH), 1-27.

Ruff, T. M. (2007). "Recommendations for Evaluating and Implementing Proximity Warning Systems on Surface Mining Equipment." National Institute for Occupational Safety and Health, 1-94.

Sacks, R., Navon, R., and Goldschmidt, E. (2003). "Building Project Model Support for Automated Labor Monitoring." *Journal of Computing in Civil Engineering*, 17(1), 19-27.

ACCELEROMTER-BASED MEASUREMENT OF CONSTRUCTION EQUIPMENT OPERATING EFFICIENCY FOR MONITORING ENVIRONMENTAL PERFORMANCE

Changbum R. Ahn[1], SangHyun Lee[2], Feniosky Peña-Mora[3]

[1] Assistant Professor, Construction Engineering and Management Division, Charles Durham School of Architectural Engineering and Construction, University of Nebraska-Lincoln, W145 Nebraska Hall, Lincoln, NE 68588; email: cahn2@unl.edu

[2] Assistant Professor, University of Michigan, 2350 Hayward St., Suite 2340 G.G. Brown Building, Ann Arbor, MI 48109; email: shdpm@umich.edu

[3] Edwin Howard Armstrong Professor of Civil Engineering and Engineering Mechanics and Professor of Earth and Environmental Engineering, and of Computer Science, Columbia University University in the City of New York, 510 Southwest Mudd Building, 500 West 120th Street, New York, NY 10027; email:feniosky@columbia.edu

ABSTRACT

Monitoring operational efficiency of construction equipment, which indicates how efficiently construction equipment is utilized, provides key information in reducing air pollutant emissions from equipment use as well as improving the productivity of construction operations. In this paper, we report our efforts to measure the operational efficiency of construction equipment, using low-cost accelerometers. The measurement of the operational efficiency of construction equipment is formulated as a problem that classifies second-by-second equipment activity into working, idling, and engine-off modes. We extract various features from the raw accelerometer data and classify them into three different equipment activities (working, idling, and engine-off), using supervised learning algorithms such as Logical Regression, decision trees, k-Nearest Neighbor, and Naïve Bayes. The result from the real-world experiment indicates that the use of supervised learning algorithms provides over 93 % of recognition accuracies and this level of accuracies causes less than 2 % error in the measurement of equipment operating efficiency.

INTRODUCTION

Tracking construction equipment operation provides important information in measuring construction productivity, analyzing activity sequence, and preventing equipment-related safety accidents. While traditional manual time studies are not reliable as well as being labor intensive , state-of-the-art tracking technologies such as Global Positioning Systems (GPS) and Radio Frequency Identification (RFID) have been used for automated location and tracking of construction equipment (Pradhananga and Teizer 2013; Wu et al. 2010; El-Omari and Moselhi

2009). In addition, vision-based techniques that analyze images from a video recording have been investigated as an alternative to GPS and RFID (Park et al. 2012; Heydarian et al. 2012; Gong and Caldas 2011).

Recently, tracking of construction equipment use has attracted a particular interest related to the environmental sustainability of construction operations because on-site construction equipment operation generates a significant amount of air pollutant emissions including carbon dioxide, nitrogen oxide, and particulate matter (EPA 2008; EPA CAAAC 2006). Reducing exhaust emissions from construction projects can be achieved with not only the use of cleaner equipment and fuel, but also the improvement of operational efficiency, which means the reduction of engine idling (Ahn and Lee 2013). Considering the great variability of Operating Equipment Efficiency (OEE), which represents the ratio of valuable (non-idle) operating time to total operating time, the improvement of OEE through detecting non-productive equipment operating (idling) is the most cost effective strategy contractors can adopt during the construction stage.

The measurement of OEEs requires activity recognition of construction equipment operation, in particular distinguishing equipment idling from other activities (e.g. working and engine-off). However, current tracking technologies such as GPS+RFID and vision-based techniques have some limitations in recognizing equipment activities compared to their capabilities in the automated localization of resources. In this context, Ahn et al. (2012) proposed the application of microelectromechanical (MEMS) accelerometers to measure OEEs of construction equipment and demonstrated its feasibility as a low-cost and nonintrusive monitoring measure. This paper extends the earlier work and proposes a systematic methodology for recognizing equipment activity by using various supervised classifiers. The paper begins with a review of the state of knowledge in the areas of automated tracking of construction equipment. Then, the research objectives and the proposed methodology are discussed. The remaining sections of the paper describe the preliminary experimental results.

BACKGROUND

Environmental Performance Monitoring

Emerging emission monitoring technologies, such as Portable Emission Measurement Systems (PEMS), allows direct measurement of air emissions from construction equipment (Lewis et al. 2009). However, such direct measurement means are too expensive for the full implementation in a project, so their use is limited to the investigation of emission rates of construction equipment for research purposes. As an alternative, the indirect measurement of construction emissions can be achieved through tracking of fuel consumption or equipment operation. The tracked fuel consumption information or equipment operation information (usage hours and OEEs) can be converted into emission amounts using pre-defined emission factors of each piece of equipment. While the collection of disaggregated fuel bills is very challenging under current construction practices, the quantification of total emission amounts using fuel bills does not guarantee the identification of mitigation opportunities for construction emissions. To identify causes of excessive idling emissions, the measurement of OEEs is critical (Ahn and Lee 2013).

However, OEEs are not tracked in current practices of construction operations. While the daily reports on the use of equipment provide information on the equipment utilized (Ahn et al. 2011), these reports do not provide information on the usage hours and OEE of each piece of equipment. This lack of disaggregated data also hinders the accurate monitoring of environmental performance.

Automated Construction Resource Tracking

Vehicle Health Monitoring Systems (VHMS) of construction equipment are available in recently manufactured models of construction equipment, and allow for the conversion of the electronic control unit (ECU) data from on-board diagnostics (OBD) protocols (e.g., OBD-II, EOBD, JOBD, and CAN bus) of construction equipment into useful information (e.g. usage, OEE, and fuel consumption). However, old equipment—which is the majority of equipment in use—does not have OBD supports, and even a piece of equipment with OBD supports requires extensive modification or the installation of additional devices. In addition, there is a compatibility issue between different manufacturers due to a lack of standardized protocol, and this creates another challenge in performing comprehensive monitoring for contractors who own, rent, or purchase machines from diverse manufacturers (Azar and McCabe 2012).

Recently, several research studies have focused on an automated tracking of construction equipment operation. Pradhananga and Teizer (2013), Wu et al. (2010), and El-Omari and Moselhi (2009) utilized GPS technology combined with RFID or wireless sensor network for the automated location and tracking of construction equipment. However, the GPS-based approach is still incapable of tracking the stationary operation of construction equipment, while it provides a high accuracy in locating resources. On the other hand, Park et al. (2012), Heydarian et al. (2012), and Gong and Caldas (2011) introduced vision-based techniques that track construction equipment operation by using video cameras. However, vision-based techniques provide only limited information based on the visual scope of cameras, and still have difficulties in providing high accuracies when there are high levels of noise (e.g., entity overlaps, moving backgrounds, and varying light conditions), which are common in dynamic construction surroundings. In addition, vision-based techniques have low performances in recognizing equipment activity, in case that target activities do not have any notable visual differences. For example, idling and some working actions of equipment (e.g. excavators' drilling dewatering holes) do now show any visual motion compared to the engine-off mode of equipment, although they create a certain level of vibration and noise.

RESEARCH OBJECTIVES AND METHODOLOGY

Accelerometers have been used in the detection of passenger car driving style (Johnson and Trivedi 2011) and the activity recognition of construction workers (Joshua and Varghese 2011). In the previous work (Ahn et al. 2012), we have attempted to detect equipment idling in a real-world operation of an excavator, using a threshold value of the acceleration signal energy. While it demonstrated the feasibility of the application of accelerometers for measuring OEEs, its determination process of a threshold value involves the subjectivity and its

detection accuracy was not that high. In this context, this paper proposes a methodology that classifies equipment operation into working, idling and engine-off modes, using supervised machine learning classification based on multiple features extracted from acceleration signals.

An overview of our approach is presented in Figure 1. Data is collected from a three-axes accelerometer mounted inside the cabin of a vehicle. The second-by-second operation of the vehicle under the study is labeled using a video recording. The accelerometer data are segmented into 128-sample windows with 50% overlapping between consecutive windows. Since the sampling rate of the experiment was 100 Hz, the window size of 128 samples is chosen as a minimum size that allows the extraction of frequency-domain features using the fast Fourier transform with the use of second-by-second activity labels. Then, a total of twenty one features are generated from each data segment, although these are all variants of just seven basic features. The computed features are average resultant acceleration (the net acceleration value), mean (for each axis), standard deviation (for each axis), peak value (for each axis), correlation (for each pair of axes), spectral entropy (for each axis), and spectral centroid (for each axis). Spectral entropy and spectral centroid are frequency domain features that represent information entropy and the weighted mean of the frequencies present in the signal. We used four different classification techniques from the the Waikato Environment for Knowledge Analysis (WEKA) Machine learning Algorithms Toolkit (Hall et al. 2009), in order to predict the activity modes of vehicles: Logical Regression, k-Nearest Neighbor (kNN), Decision Tree (J48), and Naïve Bayes. The accuracies of each classifier are evaluated using ten trials of ten-fold cross validation.
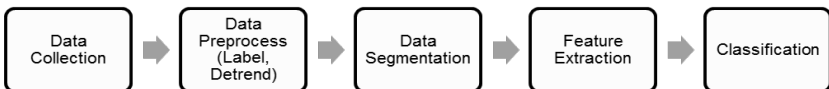
Figure 1. An overview of the proposed approach

EXPERIMENT

This section describes the experiment for the OEE measurement as well as its results.

Description of Experiment

A medium-sized wheeled excavator that performs clearing wastes on a jobsite was chosen for the experiment. The accelerometer used was the one embedded in an iPhone 4 (STMicroelectronics LIS331DLH). The sensitivity of the accelerometer was 16.2 mg (milli-g)/digit, and its measurement range was ±2 g. The sample rate of the signals is set to 100 Hz, but its actual sample rate varies between 90 and 100 Hz due to the fact that it is not a real-time system. The iPhone was installed on a rigid block near the driver seat within the cabin. The data was collected approximately 30 minutes, and the performed work of the excavator included various activities such as lifting garbage trailers, grubbing debris, and relocating (moving). The excavator was frequently idling between working cycles and its

total idling time was approximately 9.4 minutes during 30 minutes of the operation. During the last 11 seconds of the experiment, the excavator was shut off. Its OEE was measured as 68.5%. Figure 2 shows the x-, y-, and z-axis readings during the entire period of the experiment.
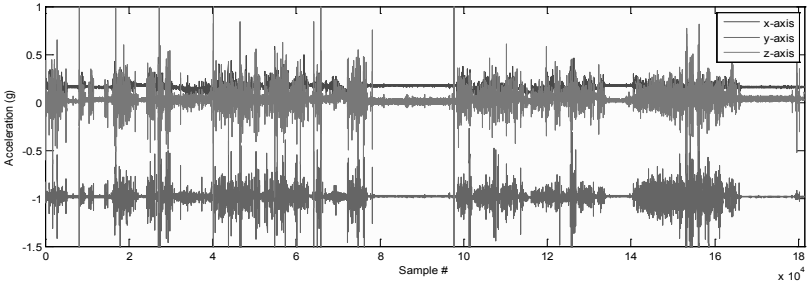
Figure 2. Raw accelerometer data collected in the experiment

Results

Once the data set was prepared, various features (13 features in time domain and 6 features in frequency domain) were computed. Figure 3 illustrates a scatterplot that depicts the values of several features: average resultant acceleration (ARA), x-axis standard deviation, z-axis peak value, and y-axis spectral entropy. This plot indicates that time domain features such as ARA, standard deviation and peak provide a good basis for classifying the activity modes of data segments. Also, frequency-domain feature, spectral entropy, provides a strong distinguishable power.
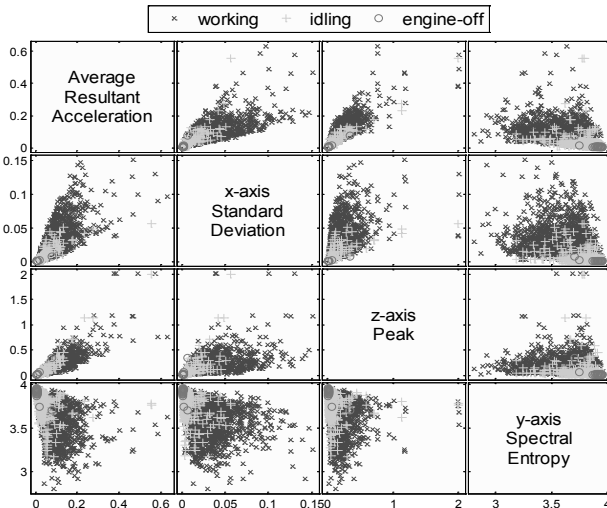
Figure 3. Scatterplot of various features extracted from raw signals

The classification of activity modes on these features was performed using various classifiers. The results are presented in Table 1. This table specifies the predictive accuracy associated with each of the four learning algorithms, the OEE measured based the predicted activities, and the deviation between actual and measured OEEs. The classification accuracies of Logical Regression, kNN, and J48 learning algorithms are over 93%, while the accuracy of Naïve Bayes is lower than other learning algorithms. The paired T-test result indicates that there is no significant difference between the performances of Logical Regression, kNN, and J48. This level of accuracies obtained from Logical Regression, kNN, and J48 creates less than ±2 percent deviations between actual and measured OEEs. Since false idling and false working in the confusion matrices offset each other in the OEE measurement, the accuracy in the OEE measurement is usually higher than the classification accuracy.

Table 1. The summery of the result

Classifiers	% of Instances Correctly Predicted (average ± standard deviation)	Measured OEE	Deviation
Logical Regression	93.85 ± 1.31	67.2%	-1.3
k-Nearest Neighbor	93.94 ± 1.41	68.2%	-0.3
Decision Tree (J48)	94.74 ± 1.29	70.1%	+1.6
Naïve Bayes	81.73 ± 2.84	62.2%	-6.3

More detailed results are presented in Figure 2 and 3, which show the confusion matrices associated with each of the four learning algorithms. While the sample size of engine-off activity is not enough to evaluate the performance of classifiers, the confusion matrices indicate that Logical Regression provides a slightly better performance in detecting engine-off and idling actions from the excavator's overall operation.

Table 2. (left) Confusion Matrix for Logical Regression; (right) Confusion Matrix for k-Nearest Neighbor (kNN)

		Predicted		
		off	idling	working
Actual	off	16	1	0
	idling	4	816	66
	working	4	104	1821

		Predicted		
		off	idling	working
Actual	off	12	5	0
	idling	6	802	78
	working	1	86	1842

Table 3 (left) Confusion Matrix for Decision Tree (J48); (right) Confusion Matrix for Naïve Bayes

		Predicted		
		off	idling	working
Actual	off	14	2	1
	idling	1	795	90
	working	1	49	1879

		Predicted		
		off	idling	working
Actual	off	16	1	0
	idling	134	692	60
	working	1	320	1609

CONCLUSIONS

This paper reports our on-going efforts to measure the operational efficiency of construction equipment using low-cost accelerometers. The experimental results show that the use of supervised learning algorithm based on multiple features extracted from acceleration signals provides over 93 % of recognition accuracies and this level of accuracies causes less than 2 % error in the OEE measurement. Compared to other tracking technologies (e.g. GPS+RFID, vision-based), the application of the proposed approach is capable of detecting a subtle change of stationary actions of equipment. In addition, the proposed approach requires much less cost and effort in the implementation, since MEMS accelerometers are now quite prevalent and cheap. Since the proposed approach is a non-intrusive measure, it can be implemented with any equipment, regardless of age or manufacturer. Considering such advantages of the proposed approach, it could be a complement to other tracking technologies that are capable of automated localization of equipment. The future work will focus on recognizing more diverse actions (e.g. left/right turn, swing arms) of construction equipment to provide detailed information for productivity analysis and safety management.

ACKNOWLEDGEMENT

REFERENCES

Ahn, C. R., Lee, S.H., and Peña-Mora, F. (2012). "Monitoring System for Operational Efficiency and Environmental Performance of Construction Operations, Using Vibration Signal Analysis." *Construction Research Congress 2012*, West Lafayette, Indiana, May 21–23.

Ahn, C., and Lee, S.H. (2013). "Importance of Operational Efficiency to Improve Environmental Performance of Construction Operations." Journal of Construction Engineering and Management, 139(4). (In press)

Ahn, C., Lee, S.H., and Pena-Mora, F. (2011). "Carbon Emissions Quantification and Verification Strategies for Large-scale Construction projects." *In Proceedings of International Conference on Sustainable Design and Construction (ICSDC)*, Kansas City, Missouri, USA, March 23–25, 2011

Azar, E. R., and McCabe, B. M. (2012). "Automated Visual Recognition of Dump Trucks in Construction Videos." J. Comput. Civ. Eng., 26(6), 768–781.

El-Omari, S., and Moselhi, O. (2009). "Integrating automated data acquisition technologies for progress reporting of construction projects." Proc., 26th ISARC. Austin, TX.

Environmental Protection Agency (EPA). (2008). *Quantifying greenhouse gas emissions in key industrial sectors*, EPA 100-R-08-002, Sector Strategies Division, US EPA, Washington D.C.

EPA Clean Air Act Advisory Committee (CAAAC) (2006). *Recommendations for reducing emissions from the legacy diesel fleet*. US EPA, Washington D.C. <http://www.epa.gov/diesel/documents/caaac-apr06.pdf> (July 24, 2010).

Gong, J., and Caldas, C. H. (2011). "An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations." *Automation in Construction*, 20(8), 1211-1226.

Hall, M., Frank, E., Holmes, G., Pfahringer, G., Reutemann, P., and Witten, I. A. (2009). "The WEKA Data Mining Software: An Update." *SIGKDD Explorations*, 11(1), 10–18.

Heydarian, A., Golparvar-Fard, M., and Niebles, J. C. (2012a). "Automated visual recognition of construction equipment actions using spatio-temporal features and multiple binary support vector machines." Proc., Construction Research Congress, West Lafayette, IN.

Johnson, D. A., and Trivedi, M. M. "Driving style recognition using a smartphone as a sensor platform." *Proc., Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, IEEE, 1609-1615.

Joshua, L. and Varghese, K. (2011). "Accelerometer-Based Activity Recognition in Construction." *J. Comput. Civ. Eng.*, 25(5), 370–379.

Lewis, P., Frey, H. C., and Rasdorf, W. (2009). "Development and Use of Emissions Inventories for Construction Vehicles," Transportation Research Record: Journal of the Transportation Research Board, National Research Council, Washington, D.C., Number 2123, 46–53.

Park, M.W. and Brilakis, I. (2012). "Enhancement of Construction Equipment Detection in Video Frames by Combining with tracking." *International Workshop on Computing in Civil Engineering*, Clearwater Beach, FL, June 17-20.

Pradhananga, N. and Teizer, J. (2013). "Automatic spatio-temporal analysis of construction site equipment operations using GPS data", Automation in Construction, 29, p.107-122.

Wu, W., Yang, H., Chew, D. A.S., Yang, S.H., Gibb, A. G.F., and Li, Q., "Towards an Autonomous Real-time Tracking system of Near-miss Accidents on Construction Sites", Automation in Construction, 19, p.134-141.

Evaluating the Impact of Location-aware Sensor Data Imperfections on Autonomous Jobsite Safety Monitoring

Xiaowei Luo, William J. O'Brien and Fernanda Leite
Department of Civil, Architectural and Environmental Engineering,
The University of Texas at Austin, Austin, TX 78712

INTRODUCTION

Construction is one of the industries with the highest fatality, non-fatal injuries and illness rates across the United States (CPWR, 2008). Although construction professionals have made great efforts to improve safety training, site supervision and design for safety (among others), the safety performance in construction is still not satisfactory due to human errors and lack of situational awareness during construction activities. Recent advances in sensing and computing technologies offer a solution for improving safety performance by providing rich information about location and hence worker safety. Our research envisions an autonomous jobsite safety monitoring system, which utilizes distributed data and information collected from various sources (RFID, localization sensors, accelerometers, load cell sensors and building information models). The research is divided into three phases: 1) knowledge elicitation,; 2) development and deployment of the autonomous safety monitoring system in a distributed computing environment; and 3) exploring the impacts of data imperfections situations (e.g., erroneous or missing data) on the safety monitoring system and approaches to reduce the impacts of the imperfections on the safety system. Prior research has reported on the first two phases (Julien, et al., 2005, Luo, et al., 2011, Luo, et al., 2011, O'Brien, et al., 2008); this paper reports the preliminary results of evaluating the impacts of data imperfections on autonomous safety monitoring using data from location-aware sensors, which provides the most important information for the autonomous jobsite safety monitoring system. The rest of this paper is organized as follows: The Related Work section summarizes the location-aware jobsite safety monitoring applications in construction and the imperfect data collected from sensors. The Research Approach section describes the research approach, including the testbed used for data collection, the test case used to evaluate the impacts of imperfect data on autonomous safety monitoring and the process of how the data is used to evaluate the impacts. The Preliminary Results section reports the preliminary results of what the impacts on autonomous safety monitoring are. Finally, the Conclusion section summarizes the results and proposes future research suggestions.

RELATED WORK

Safety at construction jobsites is a top priority for all workers. To that end, sensing technologies have been actively introduced and applied in safety applications on construction jobsites. One of the most important pieces of information used for construction safety applications is location data. There have been several different technologies available to obtain location information, including GPS, Radio Frequency (RF) sensors and video/image-based tracking. The performances of RF-based sensing technologies using different algorithms and devices for indoor and outdoor applications have been studied (Cheng, et al., 2011, Li and Becerik-Gerber, 2011, Luo, et al., 2011), and there is much space for improving their performance in terms of applications for safety monitoring.

Laser scanning has been proposed for deployment on vehicles to detect geometry and location information of the surrounding objects to prevent collisions (Riaz, et al., 2006). Similarly, ultra-wideband (UWB) for resource tracking on construction jobsites can be applied to safety applications (Teizer, et al., 2007). UWB sensors were used for implementing a proactive safety monitoring system on jobsites (Carbonari, et al., 2011). Lu (2011) described the potential uses of RFID on jobsite safety monitoring. RFIDs are used to prevent crane collisions on jobsites (Hwang, 2012) and various field sensors are used to collect crane data such as angels and lengths , and then transform some of these data into location information to monitor safe crane operations (Lee, et al., 2012, Li and Liu, 2012).

Data imperfections have been observed on sensor applications for construction jobsites. Razavi and Hass (2010) described the data imperfections from analog-to-digital converter (ADC) techniques as uncertain, incomplete, imprecise, inconsistent and ambiguous. Chi and Caldas (2012) developed a prototype simulation testbed to assist safety planners in understanding construction workspaces and accessing errors of different image-based technologies for decision making in safety planning. .

Although these studies have addressed the existing data imperfections and their impacts on construction safety decision making, they still remain in an initial phase and do not discuss the following question: what are the quantitative impacts of data imperfections on the safety monitoring decision making? This paper will address this question to help other researchers understand the requirements for sensor hardware used for jobsite safety monitoring and find other potential solutions to minimize these impacts.

RESEARCH APPROACH

To study the location-aware sensor data imperfections on autonomous jobsite safety monitoring, we choose crane safety monitoring. The research process is described in figure 1 below. First, the authors went to a construction project consisting of two office buildings with two tower cranes in the South West region of the United States. Based on the observed crane operations on the jobsite, five representative activities of tower crane swinging loads were selected to develop the test cases. In the next step, the critical elements on the jobsite involved in the

autonomous safety monitoring system were represented in a scaled test bed in a controlled lab environment. The testbed is built using LEGO models and mounted with the Cricket indoor localization system. Based on the five test cases, the discrete test points in each test case were set out in the testbed. Next, the developed crane safety monitoring system was used to collect the crane load and the worker localization data at the set test points. The location data was stored in the text files for data processing. The collected location data was then fed into the developed crane safety monitoring system as stream data to calculate the safety situations. The calculated safety situations were used to compare with the ground true safety situations to evaluate the system's performance.
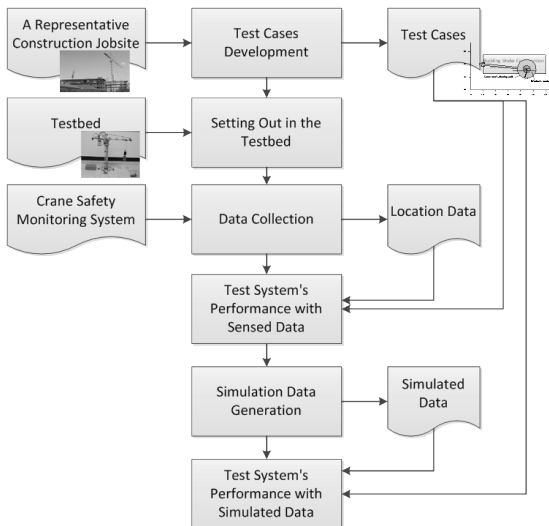
Figure 1. Research Approach Flowchart

For the full scale crane safety monitoring system, global positioning system (GPS) is a feasible and mature technology to provide location information for safety monitoring. Previous studies (Weih, et al., 2009) indicated that the localization error of GPS was usually in the range of 1 to 13 meters. Due to safety issues and difficulties to access to tower cranes, the authors decided to build a scaled testbed in a lab environment to study autonomous crane safety monitoring. The Cricket system, based on ultrasonic and radiofrequency, was selected to be mounted in the testbed to provide location information. Previous research (Priyantha, et al., 2000, Zhou, et al., 2009) indicated that the localization error of the cricket system ranges from 1 to 10 centimetres. To better evaluate the location-aware sensor data imperfections on autonomous safety monitoring, the testbed is built using the scale (100:1) between the GPS's error and the cricket system's error. The testbed is built with LEGO® blocks (Figure 2) to represent the tower crane on construction jobsites. Since the buildings, the power lines and the transportation roads have fixed locations and dimensions, it may not be necessary to represent them in the physical testbed and they can be

represented in the system by creating instances for representations. Workers are moving around the jobsite and cricket motes are used in the testbed to represent the workers and provide their location information.
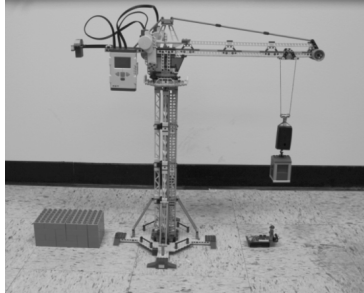
Figure 2. Scaled Testbed in the Lab Environment

The authors observed cranes' operating activities on the construction jobsite and summarized five representative test cases based on the observation. Each test case represents one crane operating activity on the jobsite and figure 3 shows test case #1 as an example. In test case #1, a worker is working right next the south side on the roof of a building under construction. The crane is clockwise swinging a steel beam over the building under construction to another building in the south. Eight discrete points were taken from the crane load's moving path. At each discrete point, twenty readings were collected from the cricket sensor and recorded into a text file. The same process was applied to the other four test cases to collect location data from the cricket sensor.
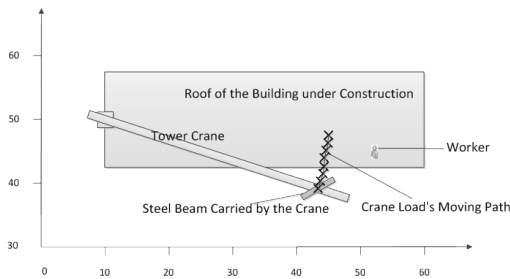
Figure 3. Test case #1

Using the crane load's location information, areas on the ground around the crane load can be divided into three areas (figure 4): 1) Red Zone, the direct area under the crane load, defined as a circle centered at the hook's location with a radius of the farthest horizontal distance between any point of the suspended load and the hook's location. No worker should stand under this area. 2) Yellow Zone, the possible fall zone under a crane load. It is an annular region between the red zone and the concentric circle with a radius of a value of the red zone's radius plus a clearance distance. Usually, 1.52 to 3.05 meters is used as the clearance. Only authorized workers can access this area. 3) Green zone: the areas outside the red zone and the

yellow zone, and the workers are unlikely to be exposed to falling load hazards in this area. A three risk levels crane safety monitoring system (Red-Yellow-Green system) contains three levels of risks, and gives warnings to the worker when he/she is in the yellow zone and gives alerts to the worker when he/she is in the red zone. To make the worker more attentive to potential hazards, the red and yellow zones can be combined, so the workers would pay the same amount of attention when they are in the yellow and red zones. This way, the system is extended into two derivatives: 1) System I (Red-Yellow-Green): crane safety monitoring with both warnings and alerts; and 2) System II (Yellow-Green): crane safety monitoring system with warnings only. In system I, an alert is triggered if the target is in the red zone and a warning is triggered if the target is in the yellow zone. In system II, a warning is triggered if the target is in the yellow zone or the red zone.
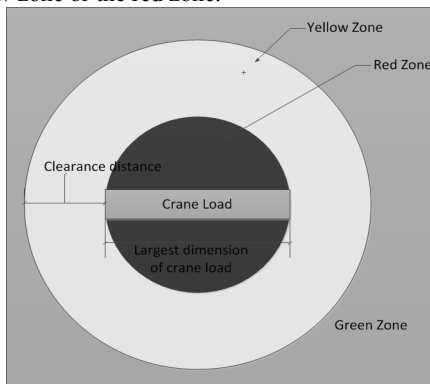
Figure 4. Dangerous Areas under the Crane Load

For each test point along the crane load's/worker's moving path in each test case, we had the sensed location data and the ground true location data. At each test point, the ground true location data is fed into the safety monitoring decision making mechanism to determine the worker's/crane's safety situations (ground true safety situation), while each sensed data reading was fed into the same mechanism to determine the safety situations (measured situation based on sensor data). At a test point at a specific timestamp, if the measured safety situation based on sensor data is not the same as the ground true safety situation, an error is detected. There are two types of errors: false positive and false negative. In a false positive error, the worker/crane is incorrectly identified to be in a more dangerous situation; in contrast with a false negative error, the worker/crane is incorrectly identified to be in a safer situation. If the measured safety situation based on sensor data is the same as the ground true safety situation, which means the worker/crane is correctly identified in safe/unsafe situations, it is true positive. The types of errors are shown in table 1.

Precision and recall are used as metrics to evaluate the impacts of location-aware sensor data on autonomous safety monitoring. Precision is the ratio of True Positive to the sum of True Positive and False Positive. It measures the percentage of the actual alerts and warnings that are triggered by the system.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad (1)$$

Recall is the ratio of True Positive to the sum of True positive and False Negative. It measures the percentage of the actual dangerous situations that are reported (when an alert or alarm is triggered by the system).

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (2)$$

If all location data is perfect (the same as ground true location data), precision and recall should be 100%. By comparing the precision and recall rates using the safety decision made upon sensor data to 100%, the impacts of sensor data imperfections can be obtained.

Table 1. Category of error in safety monitoring system

Ground True Safety Situation / Measured situation based on sensor data		Green	Yellow	Red
RYG System	Green	True Positive	False Positive	False Positive
	Yellow	False Negative	True Positive	False Positive
	Red	False Negative	False Negative	True Positive
YG System	Green	True Positive	False Positive	False Positive
	Yellow	False Negative	True Positive	True Positive
	Red	False Negative	True Positive	True Positive

PRELIMINARY RESULTS

By feeding the collected sensor data into the safety monitoring system to calculate the worker's safety situations, the corresponding precision and recall can reflect the impacts of a real combination of data imperfections on safety monitoring. Figure 5 indicates that in the RYG system, if the worker is near the center of the red zone or in the green zone, data imperfections tend to have fewer impacts on autonomous safety monitoring. In contrast, sensor data imperfections tend to have greater impacts on precision and recall near the boundary of the red zone and the yellow zone. Among the eight points in test case #1, the precision and recall rates are not satisfactory at five points. By switching the system from the RYG system to the YG system, the performance is improved. To obtain a satisfactory performance for the autonomous safety monitoring system, either improvement on sensor devices is required to reduce data imperfections, or techniques to minimize the impacts of the sensor data imperfections under the existing imperfection situations is necessary.

To further study the impacts of data imperfections (especially inaccuracy) on the safety monitoring system, the authors used simulation to explore the impacts in a controlled environment. The authors developed a set of simulated location data at each test point in the five test cases by attaching a random number following the normal distribution (to simulate the sensor error) to the ground true localization information, and used the new data set to feed into the RYG and the YG systems to evaluate the systems' performances. The results based on simulation data indicate that:

1) if the standard deviation of localization error increases, the warning systems' performances (precision and recall) degrade; 2) the performance improves by switching the system from the red-yellow-green system to the yellow-green system; and 3) for the yellow-green system, the closer a point is to the transition area of the yellow zone and the green zone, the worse the performance is.

(a) RYG System	(b) YG System
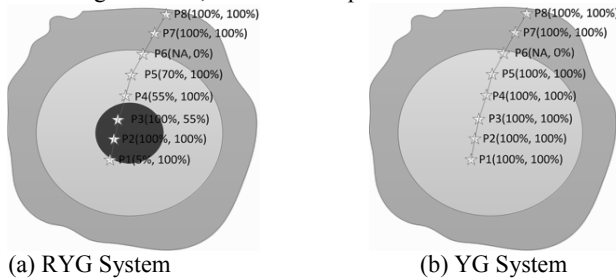
Figure 5. Precision and recall rate along the worker's moving path in test case #1

CONCLUSIONS

In this paper, the authors evaluated the data imperfections' impacts on autonomous safety monitoring using a concrete example of tower crane operations. The results indicate that the imperfections of location data collected from exiting location-aware sensors have great impacts on autonomous safety monitoring with the observed precision and recall rates of 0% to 70%. To improve the precision and recall rates of the autonomous safety monitoring system, there are some possible solutions: 1) improvement on sensor hardware to provide more perfect location information; 2) introducing approaches (e.g., Bayesian approach, data fusion from various sources) to deal with imperfect data to reduce the impacts of sensor data imperfections on safety monitoring. Future research would focus on exploring approaches to reduce the impacts of data imperfections on safety monitoring under the current level of data imperfections.

REFERENCES

Carbonari, A., Giretti, A., and Naticchia, B. (2011). "A proactive system for real-time safety management in construction sites." *Automation in Construction*, 20(6), 686-698.

Cheng, T., Venugopal, M., Teizer, J., and Vela, P. A. (2011). "Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments." *Automation in Construction*, 20(8), 1173-1184.

Chi, S., and Caldas, C. (2012). "Image-Based Safety Assessment: Automated Spatial Safety Risk Identification of Earthmoving and Surface Mining Activities." *Journal of Construction Engineering and Management*, 138(3), 341-351.

CPWR (2008). "Construction Death and Injury Rates in Selected Industrial Countries." *The Construction Chart Book: The U.S. Construction Industry and its Worker*, CPWR - The Center for Construction Research and Training, 156.

Hwang, S. (2012). "Ultra-wide band technology experiments for real-time prevention of tower crane collisions." *Automation in Construction*, 22(0), 545-553.

Julien, C., Hammer, J., and O'Brien, W. J. "A dynamic architecture for lightweight decision support in mobile sensor networks." *Proc., roc. of the Wkshp. on Building Software for Pervasive Comp*.

Lee, G., Cho, J., Ham, S., Lee, T., Lee, G., Yun, S.-H., and Yang, H.-J. (2012). "A BIM- and sensor-based tower crane navigation system for blind lifts." *Automation in Construction*, 26(0), 1-10.

Li, N., and Becerik-Gerber, B. (2011). "Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment." *Advanced Engineering Informatics*, 25(3), 535-546.

Li, Y., and Liu, C. (2012). "Integrating field data and 3D simulation for tower crane activity monitoring and alarming." *Automation in Construction*, 27(0), 111-119.

Lu, W., Huang, G. Q., and Li, H. (2011). "Scenarios for applying RFID technology in construction project management." *Automation in Construction*, 20(2), 101-106.

Luo, X., Leite, F., and O'Brien, W. J. "Requirements for Autonomous Crane Safety Monitoring." *Proc., ASCE Computing in Civil Engineering*, 331-338.

Luo, X., O'Brien, W. J., and Leite, F. (2011). "What is a safe working zone? Designing proactive fall prevention for autonomous monitoring." *2011 eg-ice Workshop*Enschede, The Netherlands, 9 pages.

Luo, X., O'Brien, W. J., and Julien, C. L. (2011). "Comparative evaluation of Received Signal-Strength Index (RSSI) based indoor localization techniques for construction jobsites." *Advanced Engineering Informatics*, 25(2), 355-363.

O'Brien, W. J., Julien, C., Kabadayi, S., Luo, X., and Hammer, J. (2008). "An Architecture for Decision Support in Ad Hoc Sensor Networks." *Journal of Information Technology in Construction*, 14, 309-327.

Priyantha, N. B., Chakraborty, A., and Balakrishnan, H. (2000). "The Cricket location-support system." *Proceedings of the 6th annual international conference on Mobile computing and networking*, ACM, Boston, Massachusetts, USA, 32-43.

Razavi, S. N., and Haas, C. T. (2010). "Multisensor data fusion for on-site materials tracking in construction." *Automation in Construction*, 19(8), 1037-1046.

Riaz, Z., Edwards, D. J., and Thorpe, A. (2006). "SightSafety: A hybrid information and communication technology system for reducing vehicle/pedestrian collisions." *Automation in Construction*, 15(6), 719-728.

Teizer, J., Lao, D., and Sofer, M. (2007). "Rapid Automated Monitoring of Construction Site Activities Using Ultra-Wide Bandq." *24th International Symposium on Automation & Robotics in Construction (ISARC 2007)*Kochi, India, 23-28.

Weih, R. C., Gilbert, M., Cross, J., and Freeman, D. (2009). "Accuracy Assessment of Recreational and Mapping Grade GPS receivers." *Journal of the Arkansas Academy of Science*, 63, 163-168.

Zhou, S., Feng, H., and Yuan, R. "Error Compensation for Cricket Indoor Location System." *Proc., Parallel and Distributed Computing, Applications and Technologies, 2009 International Conference on*, 390-395.

Monocular Videogrammetry for Generating Absolute-Scale 3D Point Cloud Data of Civil and Infrastructure Scenes

Rashidi[1], I. Brilakis[2] and P. Vela[3]

[1]PhD Candidate, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332; email: rashidi@gatech.edu
[2]Laing O'Rourke Lecturer of Construction Engineering, Department of Engineering, University of Cambridge, UK; email: ib340@cam.ac.uk
[3]Associate Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332; email: pvela@gatech.edu

ABSTRACT

Existing methods for automatically determining the absolute scale of Point Cloud Data (PCD) acquired from monocular video/photogrammetry are suitable to a limited set of applications under specific settings and are not general enough to be considered to be practical solutions for reconstructing both indoor and outdoor built infrastructure scenes. To address the issue, in this paper, we propose a novel method for automatically calculating the absolute scale of built infrastructure PCD. The absolute scale estimation method uses a pre-measured cube for outdoor scenes and a letter-size sheet of paper for indoor environments as the calibration patterns. Assuming that the dimensions of these objects are known, the proposed method extracts the objects' corner points in 2D video frames using a novel algorithm. The extracted corner points are then matched between the consecutive frames. Finally, the corresponding corner points are reconstructed along with other features of the scenes to determine the real world scale. Three indoor and three outdoor cases were selected to evaluate the performance of the proposed method and the absolute-scale PCD for each case was computed. Obtained results illustrate the capacity of the proposed method to accurately compute the absolute scale of PCD.

INTRODUCTION

Reconstruction of the 3D structure of built infrastructure is mainly presented in the form of point cloud data (PCD). In civil engineering, PCD are typically utilized in 3D as-built documentation of infrastructure, quality control of construction-related products, effective progress monitoring of projects, and deviation identification of constructed facilities from as-planned conditions (Brilakis et al. 2011).

3D reconstruction is possible with both active (e.g. laser) and passive (e.g. camera) sensors (Dai et al. 2013). Within the last two decades, advances in high resolution digital photography and increased computing capacity, have made it possible for image-based 3D reconstruction to produce highly accurate results.

3D reconstruction algorithms based on video are divided into three separate categories based on the number of cameras used (Rashidi et al. 2013):

1- Monocular setting or using a single camera as the sensor

2- Binocular setting or using a stereo set of cameras as the sensor

3- Camera rigs or using multiple cameras setup as the sensor

For regular applications in the fields of construction engineering and facility management, using a single camera is the easiest, most practical way to capture images/video data. However, in computer vision it is well known that using a monocular camera only allows for the creation of unknown global scale PCD (Scaramuzza et al. 2009). In order to compute the absolute scale, the operator needs to know the base line of the camera motion or at least one dimension of the scene. Manually measuring the dimensions in job sites needs extra work and the results might be inaccurate. Further processing is also required for identifying the measured dimensions on PCD and scaling the entire scene proportionally, which is a labor-intensive task.

In this paper, we propose a novel method for automatically computing the absolute scale of PCD. The proposed method is based on using pre-measured simple objects, particularly a letter size sheet of paper for indoor settings and a simple colored cube made of plywood material for outdoor environments. The vertices of these predefined objects are detected in video frames using a novel algorithm. The detected vertices in 2D frames are then reconstructed along with the other feature points extracted from the scene. Knowing the distance between the vertices, the entire PCD is then scaled up using an existing method. The paper is organized as follows: The background section summarizes the existing research on absolute scale calculation for monocular photo/videogrammetry. Our method for automating the absolute scale calculation is presented in the next section. In Experiments' section, experiments are conducted to test the validity of the proposed algorithms and the entire pipeline. Finally, conclusions are drawn in the last section.

BACKGROUND

The standard approach for acquiring the absolute scale in 3D reconstruction is implementing a binocular setup using stereo cameras with known baseline (Brilakis et al., 2011). Binocular photo/videogrammetry requires specific setup and adjustments, e.g. aligning the lenses, which prevents its general application in routine AEC/FM practices. In comparison with stereo cameras, using a single camera (monocular setup) is more practical since there is no need for specific setup and almost everyone on a construction job site has access to an off-the-shelf digital camera or smart phone for videotaping the scene or taking pictures. In the case of using a single camera, a scene can only be reconstructed up to an unknown scale factor. This limitation is of great significance since in civil engineering and facility management applications; all measurements take place in Euclidean space with real values. Besides manually measuring one dimension of the scene, two major approaches might be used for automatically recovering the absolute scale: 1) obtain prior knowledge about the scene and existing objects (Kuhl et al. 2006; Tribou 2009) and 2) use additional sensors, e.g. accelerometers and GPS, for getting extra information about the scene or motion of the camera (Kneip, et al. 2011; N¨utzi et al. 2011 and Eudes et al. 2010).

In the area of AEC/FM, a number of researchers applied specific settings to solve particular problems. Golparvar-Fard et al. (2012) used 3D coordinates of

predominant benchmarks, e.g. corners of walls and columns, and the building information modeling (BIM) of the built infrastructure to solve the absolute scale and registration problems. Jahanshahi et al. (2011) on the other hand, applied the working distance (camera-object distance), the camera sensor size and the camera sensor resolution to estimate the dimensions of cracks while reconstructing the surfaces of structural elements.

As a major obstacle, the existing solutions are limited to specific categories of scenes and are not general enough to cover a vast range of applications for the AEC/FM practices. Moreover, the need for extra sensors and associated hardware constraints prevent some of the solutions from gaining popularity in construction and facility job sites. As a result, there is significant demand for a simple, practical solution applicable for both indoor and outdoor built infrastructure scenes. The research objective of this paper is to test whether a novel method proposed by the authors is able to successfully and accurately compute the absolute scale of various built infrastructure scenes in both indoor and outdoor environments. The key research question that will be addressed is: how can we automatically compute the absolute scale of built infrastructure scenes using simple objects easily placed in almost all job sites, and without requiring any extra sensors or hardware setup?

The presented solution relies on using predefined objects, with known dimensions, for each indoor and outdoor scenario in order to extract the necessary prior knowledge about the scene. The proposed method is described in details in the following sections:

METHODOLOGY: AUTOMATED ABSOLUTE SCALE COMPUTATION FOR OUTDOOR SETTINGS

Many AEC/FM practices take place in outdoor settings so it is necessary to choose a simple, consistent object which is easily detectable and easy to build at most job site. Among geometrical objects, a cube is the simplest. The dimensions of a cube are equal and it is typically possible to view three of its surfaces from various perspectives simultaneously. We chose a cube made of plywood which is solid and light weight, noting that it can be easily built at nearly any job site. The size of the cube should be big enough to use in large scale infrastructure scenes yet small enough to be carried out and handled by only one person. Considering those factors we choose 0.8 meter as the standard dimension for the cube .

In order to better detect the object in the scene we chose three different colors for the cube's surfaces. Two criteria should be considered while choosing the right colors for the cube surfaces: One, the colors should be distinct from the colors of existing features in the scene, and two, there should be a maximum difference between RGB (HSV) values of the selected colors so they can easily be identified using color detection algorithms. Considering the above constraints, we remove colors close to blue and green since those colors frequently appear in outdoor settings. Examining what remains, and distributing the color values as evenly as possible across the remaining spectrum, leads to the three distinct colors whose HSV values are depicted in Figure 1.

Given the selected colors, the overall method for calculating absolute scale mainly relies on detecting the cube in video key frames; identifying, matching and

reconstructing the cube vertices along with other feature points of the scene; and scaling the obtained PCD given the known dimensions of the cube (distances between the vertices). Figure 2 depicts the proposed framework for absolute scale estimation.

Figure 1: Selected colors for surfaces of the cube.
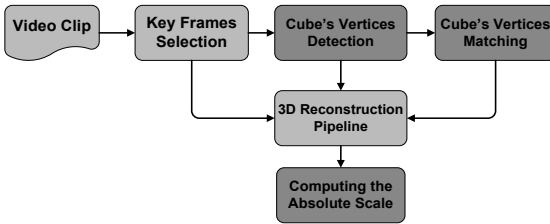
Figure 2: Overall workflow of proposed algorithm for computing absolute scale of PCD

More details about different stages of the proposed algorithm are presented in the following three steps:

Step 1: Detection of the cube's vertices: Figure 3 describes the necessary steps for detecting the vertices of the cube in 2D video frames captured from the scene.

The procedure starts with detecting the surfaces of the cube by filtering the HSV values. For each detected surface, the connected components are analyzed and an opening morphology operator is applied to remove small areas with the same color values which do not belong to the cube's surface. To ensure that detected areas belong to the cube surfaces, the following constraints should be met:

1- The area of the surface should be bigger than 0.005 of the area of the entire image. This criterion removes false detections of small areas that might match, and also ignores detected boxes that are too far from the camera which often introduce estimation error. The threshold value, 0.005, was experimentally obtained.

2- It is assumed that each surface of the cube should look neither too long nor too circular in the image. Accordingly, the roundness of the surface, calculated by:

$$Roundness = \frac{4\pi \times Area}{(Perimeter)^2} \tag{1}$$

should be located between an upper and a lower threshold.

3- Due to the perspective projection equations describing image formation, the imaged surfaces of a cube are trapezoidal in shape, which is convex. To isolate potential cubes by removing non-convex objects, the real area of the surface should be approximately equal to the convex hull of the surface computed by implementing an algorithm called the "vertex hull algorithm."
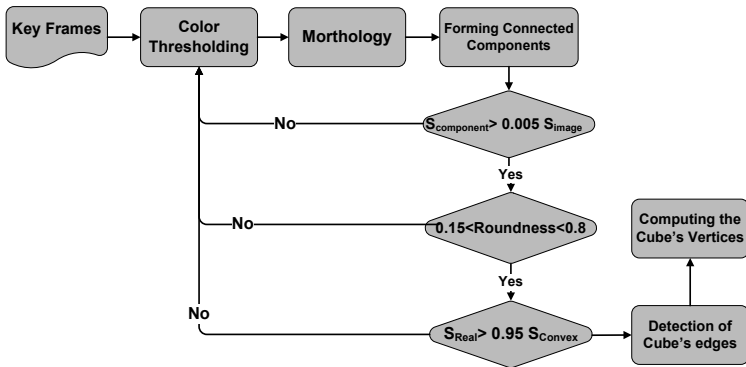
Figure 3: Necessary steps for detection of the cube vertices.

After identifying the surfaces of the cube, the edges of the cube are detected using a modified version of the Hough transform. Due to nonlinear lens distortions, the cube edges may not appear straight in the 2D images, but will be curved. The Hough transform algorithm was adjusted to also detect curved lines by adding dilation as a parameter to the search space. Finally, the cube vertices are identified by determining neighboring edges through their intersection points.

Step 2: Matching the cube's vertices across key frame views: After detection of the cube's vertices, the next step is to correspond these vertices within two key frame views. Our matching strategy consists of two components:

1-   The corresponding point for each vertex in one key frame view should be located on the epipolar line for the other view. The epipolar constraints of the scene are computed from an estimate the camera motion between the two key frames.

2-   Applying the color differences is the second criterion. We consider a rectangular window around each vertex. Since the motion of the camera between two consecutive key video frames is small, we expect that the corresponding window in the other frame also contains similar color values. In other words, the best corresponding window is selected by following a cross correlation approach and calculating the similarity score (Rashidi et al, 2011),

$$COL\_DIFF(w_1,w_2) = \sum_{j=1}^{n}\sum_{i=1}^{m}abs(I_{i,j}-I'_{i,j}) + \sum_{j=1}^{n}\sum_{i=1}^{m}abs(R_{i,j}-R'_{i,j}) + \sum_{j=1}^{n}\sum_{i=1}^{m}abs(G_{i,j}-G'_{i,j}) + \sum_{j=1}^{n}\sum_{i=1}^{m}abs(B_{i,j}-B'_{i,j}) \quad (2)$$

$$CORR(w_1,w_2) = \sum_{j=1}^{n}\sum_{i=1}^{m}I_{i,j}I'_{i,j} + \sum_{j=1}^{n}\sum_{i=1}^{m}R_{i,j}R'_{i,j} + \sum_{j=1}^{n}\sum_{i=1}^{m}G_{i,j}G'_{i,j} + \sum_{j=1}^{n}\sum_{i=1}^{m}B_{i,j}B'_{i,j} \quad (3)$$

$$SIMILARITY\_SCORE(w_1,w_2) = \frac{CORR(w_1,w_2)}{1+COL\_DIFF(w_1,w_2)} \quad (4)$$

where R, G, B and I are the individual color channel and intensity values of the neighborhood pixels of the windows constructed around each vertex.

Step 3: 3D reconstruction of the cube's vertices along with other features of the scene: After computing the corresponding vertices within the consecutive frames, we use a standard 3D reconstruction pipeline to reconstruct the vertices of the cube as

well as other features of the scene (Rashidi et al, 2011). We used the Patch-based Multi-view Stereo approach, as implemented in (Rashidi et al, 2013), to reconstruct the entire scene and obtained the PCD. Assuming that the dimensions of the cube are known, we can scale up the entire PCD. It is necessary to mention that in order to calculate the absolute scale; at least three vertices of the cube should be successfully reconstructed.

METHODOLOGY: AUTOMATED ABSOLUTE SCALE COMPUTATION FOR INDOOR SETTINGS

Our suggestion for a proper object for use in indoor settings is a simple letter-size sheet of paper. Letter-size paper can be found in almost every indoor environment, including homes and offices. The placement of the paper should on a darker uniform surface to maximize detection.

The algorithm for detecting, matching and reconstructing the corners of the sheet of the paper is the same as those of the cube with the exception of the matching stage. Since we are only dealing with four points as the corners of the paper, it suffices to implement the epipolar geometry constraint, and taking note that the four corners in the first view and their correspondences in the second view are located based on a same clockwise order.

IMPLEMENTATION AND RESULTS

A C#-based prototype was implemented to test the validity of the proposed method. Considering the variety in built infrastructure scenes, we selected three indoor and three outdoor cases as our case studies. Each scene was videotaped as completely as possible, with sensing from multiple viewpoints to minimize occlusions. The validation took place in two stages: As the first step, the proposed algorithms for detection and matching the vertices of the cube and the corners of the paper sheet were evaluated separately. As the second step, the entire scenes were reconstructed and the obtained PCD were scaled up as explained before. The average errors in measuring length of several objects of the scenes, obtained using the proposed scaling up algorithm, were also computed. Table 1 depicts the summary of the results in terms of accuracy and percentages of successful implementing the algorithms. Snapshots of the scenes and samples of computed PCD are also presented in figure 4.

Table 1: Summary of the results obtained from implementing the proposed method.

Types of error	Indoor settings	Outdoor settings
Average percentage of successfully detecting the cube/paper surfaces (%)	93.7	89.4
Average percentage of successfully detecting the cube/paper vertices (%)	98.4	96.2
Average error in the computed location of the vertices/corners in 2D frames (pixels)	2	3
Average percentage of successfully corresponded vertices in consecutive key frames (%)	94.5	87.6
Average overall error of the length measurements (error in the ratio of computed lengths to actual lengths after absolute scaling the PCD (%)	0.5	0.95

Figure 4: Snapshots of indoor (top) and outdoor (bottom) case studies and generated PCD.

SUMMARY AND CONCLUSION

This paper presented and validated an innovative method for automatically computing the absolute scale of PCD's obtained from indoor/outdoor built infrastructure scenes. Computing the absolute scale of PCD is a major issue faced by civil engineers and facility managers since they need to extract the real measurements from video-generated PCD with scale uncertainty. The proposed algorithm is based on detecting, matching, and reconstructing the corner points of two simple categories of objects: a letter size piece of paper for indoor applications and a plywood cube for outdoor, large scale cases. The average length measurement errors resulted by implementing the proposed method for indoor and outdoor scenarios were 0.95 cm and 2.21 cm respectively. The experiment results revealed that the proposed method enables AEC/FM practitioners to accurately scale up PCD with least amount of manual work and without the need for extra sensor/prior knowledge about the scene. As the extension of the current research, the authors will conduct more experiments in both indoor and outdoor settings to better evaluate the performance of the method.

REFERENCES

Brilakis, I., Fathi, H., and Rashidi, A. (2011). "Progressive 3D reconstruction of infrastructure with videogrammetry." *Automation in Construction*, 20(7), 884-895.

Dai, F., Rashidi, A., Brilakis, I., and Vela, P. (2013). "Comparison of image- and time-of-flight-based technologies for three dimensional reconstruction of infrastructure." *Journal of Construction Engineering and Management*, ASCE, 139(1), 69–79.

Eudes, A., Lhuillier, M., Naudet, S., and Dhome, M. (2010). "Fast odometry integration in local bundle adjustment-based visual SLAM." *Proc., In 20th*

International Conference of Pattern Recognition (ICPR), Istanbul, Turkey, 290-293.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012). "Automated progress monitoring using unordered daily construction photographs and IFC-based building information models." *Journal of Computing in Civil Engineering*, in press.

Jahanshahi, M. R., Masri, S. F., Padgett, C. W., and Sukhatme, G. S. (2013). "An innovative methodology for detection and quantification of cracks through incorporation of depth perception." *Machine Vision and Applications,* 24(2), 227-241.

Kneip, L., Martinelli, A., Weiss, S. Scaramuzza, D., and Siegwart, R. (2011). "Closed-form solution for absolute scale velocity determination combining inertial measurements and a single feature correspondence." *Proc., In IEEE International Conference on Robotics and Automation (ISRA),* Shanghai, China, 4546-4553.

Kuhl, A., Wöhler, C., Krüger, L., d'Angelo, P., and Groß, H. M. (2006). "Monocular 3D scene reconstruction at absolute scales by combination of geometric and real-aperture methods." *Lecture Notes in Computer Science,* 4174, 607-616.

Nützi, G., Weiss, S., Scaramuzza, D., and Siegwart, R. (2011). "Fusion of IMU and vision for absolute scale estimation in monocular SLAM." *Journal of Intelligent and Robotic Systems (JIRS)*, 61(1-4), 287-299.

Rashidi, A., Fathi, H., and Brilakis, I. (2011). "Innovative stereo vision-based approach to generate dense depth map of transportation infrastructure." *Transportation Research Record, Journal of the Transportation Research Board*, 2215, 93-99.

Rashidi, A., Dai, F., Brilakis, I. and Vela, P. (2011). "Comparison of Camera Motion Estimation Methods for 3D Reconstruction of Infrastructure", Proceedings of the 2011 ASCE International Workshop on Computing in Civil Engineering, Miami, FL

Rashidi, A., Dai, F., Brilakis, I., and Vela, P. (2013). "Optimized selection of key frames for monocular videogrammetric surveying of civil infrastructure." *Journal of Advanced Engineering Informatics*, accepted.

Scaramuzza, D. Fraundorfer, F. Pollefeys, M., and Siegwart, R. (2009). "Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints." *Proc., In IEEE International conference on computer vision,* Kyoto, 1413-1419.

Tribou, M. (2009). "Recovering scale in relative pose and target model estimation using monocular vision." A thesis presented to the University of Waterloo, Waterloo, Ontario, Canada.

Detecting, Fitting, and Classifying Surface Primitives for Infrastructure Point Cloud Data

G. Zhang[1], P. A. Vela[1] and I. Brilakis[2]

[1] School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0250; email: zhanggc@gatech.edu, pvela@gatech.edu
[2] Department of Engineering, University of Cambridge BC2-07, Trumpington Street, Cambridge, CB2 1PZ, UK; email: ib340@eng.cam.ac.uk

ABSTRACT

This paper presents a novel algorithm for detecting, fitting and classifying the embedded surface primitives from a point cloud dataset (PCD). Given a noisy infrastructure PCD the final output of the algorithm consists of segmented surfaces, their estimated quadric models and corresponding surface classification. Initially, the PCD is down-sampled with a k-d tree structure then segmented via subspace learning. After pose recovery for each segmented group via singular value decomposition, a full quadric model is fit in MLESAC using the direct linear transform for parameter estimation. From the model parameters, the surface is classified from the rank, determinant, and eigenvalues of the parameter matrices. Finally model merging is performed to simplify the results. A real-world PCD of a bridge is used to test the algorithm. The experimental validation of the algorithm demonstrates that the surface primitives are accurately estimated and classified.

INTRODUCTION

The conditions of a facility during or after construction are not always truthfully represented by the as-designed documentation. In contrast, the as-built building information models (BIMs) do describe an existing facility more accurately. A typical process to generate as-built BIMs uses raw PCD from remote-sensing or photogrammetry as input and outputs an information-rich object oriented models. However, currently generating as-built BIMs is cost-prohibitive, because the conversion from the raw PCDs to the geometric models is a very time-consuming manual process. Therefore, automating this conversion process is of great significance for greater use of as-built BIMs.

To address this problem, Zhang (Zhang, G., et al, 2012) proposed a sparsity-inducing optimization based algorithm to extract planar patches from noisy PCDs. The algorithm first segments the PCD into linear subspaces and then performs robust model estimation with planar models. This approach has been demonstrated to be effective for real-world infrastructure PCDs. However, it only works for planar surfaces. In contrast, this work seeks to extract more kinds of structures from infrastructure PCDs. This work focuses on processing PCDs instead of the registered images to make the algorithm applicable to not only image-based reconstructed PCDs but also PCDs from laser scanners. The algorithm is designed to detect, fit, and

classify multiple surface primitives robustly and efficiently for civil infrastructure PCDs. There are major advances over the previous one: (1) the segmentation of PCD is based on correlation relationships and performed in a fast manner; (2) the model estimation and the classification scheme is designed to make the algorithm work for much more geometric shape primitives; (3) a model merging procedure is performed with a novel model similarity measures to reduce the redundancy of the algorithm output. The algorithm is tested on a real-world infrastructure PCDs from photogrammetry. The results are evaluated quantitatively with two evaluation metrics.

RELATED WORK

3D object detection is well-studied in information retrieval. One of the most widely used 3D object detection methods is the graph-based method which encodes both geometric and topological information. Examples include skeletal graphs (Sundar, 2003) and augmented multi-resolution Reeb graphs (Tung, 2004). A second class of methods is based on geometry, for instance, the principal plane analysis proposed in (Kuo, 2007). A survey is presented by Tangelder (Tangelder, 2008) on 3D object classification and retrieval methods. All of these methods take queried 3D models in specific formats as input and finally output the matching class. However, they are not specially designed or optimized for PCDs and require a large database for 3D models.

Within computer vision and robotics, object extraction algorithms specially designed for PCD often utilize classification and recognition algorithms. These methods, mostly being descriptor-based, are categorized into two groups: local methods and global methods. Local methods exploit descriptors based on locally invariant geometric properties around a surface point (Johnson, 1999; Mian, 2006; Schnabel, 2007; Rusu, 2009). These methods take a PCD, extract the descriptors, and then use the extracted descriptors as input to a classifier. Local methods have the advantage of computational efficiency, but the performance heavily depends on the quality and resolution of the input PCD (Drost, 2010). Moreover, these methods do not work globally and are designed for free-form object recognition. Thus they are not suitable for solving the problem we seek to address.

Of the global methods, the two main categories are generalized Hough transforms and RANSAC. Hough transform based methods first map input PCDs to parameter spaces and then classify shapes (Rabbani, 2005; Zaharia, 2001). Nevertheless these methods are expensive in both computation and memory. For RANSAC paradigm methods, Schnabel (Schnabel, 2007) proposed an efficient variant of RANSAC to detect and recognize primitives in PCDs, but it only works for five shapes and requires different estimation for different shapes.

In order to overcome all the issues above, in this work we aim to develop a novel global method for detection, fitting, and classification of surface primitives. The method should be easily extended and general enough for a variety of primitives. Also the robustness and speed of the algorithm are important factors to be considered.

PROPOSED ALGORITHM

The proposed algorithm for detecting, fitting and classifying PCD surface primitives consists of several steps as outlined in **Figure 1**. The algorithm takes a raw PCD (with only Cartesian information) as input. Firstly, a fast segmentation is

performed on the PCD. Each segmented group of points corresponds to a surface patch. The 6-DOF pose of the segmented points is recovered by finding a $SE(3)$ matrix transforming the points to a canonical coordinate system. Then a full quadric model is fit to the segmented points using least-square estimation in a Maximum Likelihood Estimator SAmple Consensus (MLESAC) paradigm (Torr, P. H., et al, 2000). Using the estimated model parameters, the surface is classified as 1 of 12 types of quadric surfaces. After all groups have been processed, model merging is performed to join abutting groups from the same surface class, after which the models of the joined groups are re-estimated. Finally the algorithm outputs the estimated surface parametric models, surface primitive types, along with the PCD segmentation.
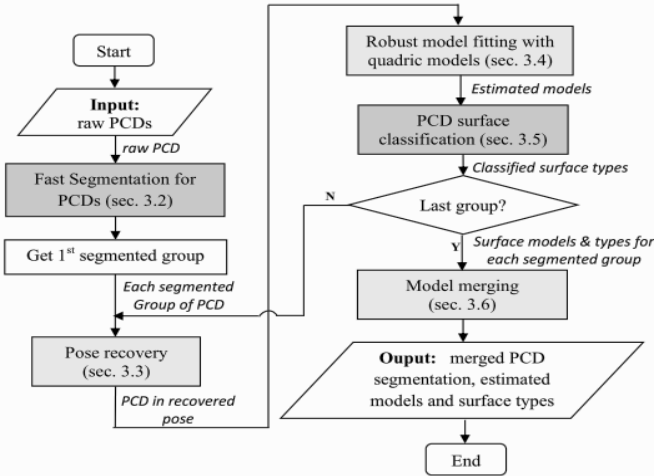
Figure 1. Block diagram of the proposed method; yellow blocks represent brand new algorithms; blue blocks represent modified algorithms from existing ones

FAST SEGMENTATION OF PCDS.

The fast segmentation strategy for large-scale PCDs down-samples the dataset, performs segmentation, and finally retrieves the labels for the original PCD. The original PCD $\mathcal{P}_0 \in \mathbb{R}^{N \times 3}$ is down-sampled by partitioning the PCD into discrete cubes in $\mathbb{R}^3$ and replacing the points in each cube with their centroid. The partitioning is done by a k-d tree search, which generates adaptive partitions and is efficient for large-scale data. A static k-d tree for a set P of n points takes $O(n \log n)$ time to build.

Segmentation is performed on the down-sampled point-set $\mathcal{P}' \in \mathbb{R}^{N' \times 3}$. The segmentation step exploits the local geometric relationship among points on the same quadric surface, i.e. a correlation relationship. The segmentation follows the idea in (Ma, Y., et al, 2007), and clusters the PCD via lossy-compression. Assume $\mathcal{P}'$ is from a mixture of Gaussians and denote a segmentation into K clusters as $\mathcal{P}' = \{\mathcal{W}_1' \cup \mathcal{W}_2' \cup \dots \cup \mathcal{W}_K'\}$, then the total number of bits to encode $\mathcal{P}'$ up to distortion λ is:

$$L^s(\mathcal{W}_1', \mathcal{W}_2', \dots \mathcal{W}_K') = \sum_{i=1}^{K} \left[L(\mathcal{W}_i') - |\mathcal{W}_i'| \log_2 \left(\frac{|\mathcal{W}_i'|}{N'} \right) \right], \quad (1)$$

for which $L(\mathcal{W}_i')$ is the number of bits needed to encode each cluster $\mathcal{W}_i'$:

$$L(\mathcal{W}_i') = \frac{N'+3}{2}\log_2 \det\left(I + \frac{3}{\lambda N'}\overline{\mathcal{W}_i'}\overline{\mathcal{W}_i'}^T\right) + \frac{3}{2}\log_2(1 + \frac{\mu_i^T\mu_i}{\lambda}), \tag{2}$$

where μ is the mean of $\mathcal{W}'$ and $\overline{\mathcal{W}'} = \mathcal{W}' - \mu$.

It is proved in (Ma, Y., et al, 2007) that $\Delta L = L\left(\mathcal{W}_i' \cup \mathcal{W}_j'\right) - \left[L\left(\mathcal{W}_i'\right) + L\left(\mathcal{W}_j'\right)\right] \geq 0$, and that when $\mathcal{W}_i'$ and $\mathcal{W}_j'$ are more correlated, then ΔL is smaller. The segmentation minimizing the coding (bit) length will segment $\mathcal{P}'$ into clusters that maximize the intra-cluster correlation and minimize the inter-cluster correlation. This optimal segmentation is found by a pairwise steepest descent procedure.

After segmenting $\mathcal{P}'$, the segmentation for $\mathcal{P}_0$ is obtained by labeling each point $\mathit{p}_{0i} \in \mathcal{P}_0$ as belonging to the cluster $\mathcal{W}_k' \in \mathcal{P}'$ with the minimum point-to-cluster distance $D_{\mathit{p}}^{\mathcal{W}_k'}$. Let $\mathit{p}_j^{\mathcal{W}_k'} \in \mathcal{W}_k'$, then the label for p_{0i} is:

$$k_{\mathit{p}_{0i}} = \mathrm{argmin}_k \ D_{\mathit{p}}^{\mathcal{W}_k'} = \mathrm{argmin}_k \left\{ \min\left\{ \sqrt{\left\|\mathit{p}_{0i} - \mathit{p}_j^{\mathcal{W}_k'}\right\|_{l_2}} \right\}_{j=1}^{|\mathcal{W}_k'|} \right\}_{k=1}^{K}. \tag{3}$$

PCDs Pose Recovery

Before model estimation, the PCD in each segmented group should be transformed to a canonical pose to improve the conditioning of the quadric model parameter estimation. To find the rigid transformation $T \in SE(3)$ for transforming the PCD $\mathcal{P} \in \mathbb{R}^{N\times 3}$, $\mathcal{P}$ is first translated by subtracting the centroid $\mathcal{C}_\mathcal{P} \in \mathbb{R}^{1\times 3}$ to obtain $\bar{\mathcal{P}} = \mathcal{P} - \mathcal{C}_\mathcal{P} \cdot I_{N\times 1}$. Then Singular Value Decomposition is performed on $\bar{\mathcal{P}}^T$:

$$\bar{\mathcal{P}}_{3\times N}^T = R_{3\times 3}\, S_{N\times N}V_{N\times 3}\,, \tag{4}$$

where $R_{3\times 3}$ is a unitary matrix which can be viewed as a matrix in $SO(3)$, and $S_{N\times N}$ is a diagonal matrix encoding the scale. The rigid transformation T is then given by:

$$T = \begin{bmatrix} R & -\mathcal{C}_\mathcal{P}^T \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \tag{5}$$

The PCD $\tilde{\mathcal{P}} \in \mathbb{R}^{N\times 3}$ in the recovered pose is obtained from:

$$\begin{bmatrix} \tilde{\mathcal{P}} \\ 1 \end{bmatrix} = \begin{bmatrix} S_{N\times N}V_{N\times 3} \\ 1 \end{bmatrix} = T \cdot \begin{bmatrix} \mathcal{P} \\ 1 \end{bmatrix}. \tag{6}$$

Robust Fitting with Quadric Models.

After pose recovery, to every PCD $\tilde{\mathcal{P}}$, a fully quadric model is fit using least-square estimation. The quadric model is:

$$\tilde{\mathcal{P}}^T\Theta\tilde{\mathcal{P}} = 0;\ \Theta \in \mathbb{R}^{10\times 10}, \tag{7}$$

which can be expanded as:

$$ax^2 + by^2 + cz^2 + fyz + gzx + hxy + px + qy + rz + d = 0. \tag{8}$$

The parameter vector $\theta = [a,b,c,f,g,h,p,q,r,d]^T$ can be estimated with direct linear transformation (DLT). Let $\tilde{\mathcal{P}} = \{[x_i,\ y_i,\ z_i]\}_{i=1}^N$, the algebraic error is

$$\varepsilon = \begin{bmatrix} x_1^2, y_1^2,\ z_1^2, y_1z_1, z_1x_1, x_1y_1, x_1, y_1, z_1, 1 \\ x_2^2,\ y_2^2,\ z_2^2, y_2z_2, z_2x_2, x_2y_2, x_2, y_2, z_2, 1 \\ \vdots \\ x_N^2,\ y_N^2,\ z_N^2, y_Nz_N, z_Nx_N, x_Ny_N, x_N, y_N, z_N, 1 \end{bmatrix}\theta \triangleq \mathbf{A}\cdot\theta \in \mathbb{R}^{N\times 1}. \tag{9}$$

The estimate is obtained by minimizing $\|\varepsilon\|^2$, which can be done by performing SVD on the data matrix $\mathbf{A}$. Let $\mathbf{A} = USV^T$, then θ is given by the tenth column of V.

The least-square estimation of θ is performed using MLESAC. MLESAC is an accurate and robust method in the presence of measurement uncertainty and noise (Choi, et al, 2009). It is also invariant to the number of points, which means that the threshold need not vary with the sample size. The inlier error e is modeled by an unbiased Gaussian distribution, while the outlier error is modeled by a uniform distribution. The loss is defined as $Loss(e) = -\ln[Prob(e|\theta)]$, where $Prob(e|\theta)$ is the probability of the error given the estimated model. If the loss of the estimated model is smaller than a threshold, then the model will be re-estimated using only the inliers and the iterations terminate; otherwise repeat the random sampling and estimation step up to a maximum number of iterations. The maximum number of iterations is:

$$T = \log \alpha / \log(1 - \gamma^{10}), \tag{10}$$

where γ is the prior probability of being an inlier, and α is the probability of failing to pick a valid inlier set during the random sampling.

CLASSIFICATION OF SURFACE PRIMITIVES.

In classification step, 12 quadric surface primitives are considered, as shown in Table 1. The estimated model parameter vector θ is first truncated by setting the elements smaller than a threshold to be zero. Then form the following two matrices:

$$e = \begin{bmatrix} a & h/2 & g/2 \\ h/2 & b & f/2 \\ g/2 & f/2 & c \end{bmatrix} \text{ and } E = \begin{bmatrix} a & h/2 & g/2 & p \\ h/2 & b & f/2 & q \\ g/2 & f/2 & c & r \\ p & q & r & d \end{bmatrix} \tag{11}$$

The classification parameters are: $\Delta = \det(E)$; $\phi = rank(e)$; $\Phi = rank(E)$; "k-sign" and "K-sign" refer to the nonzero eigenvalues of e and E respectively

Table 1. Descriptions and classification criterions for quadric surface primitives

No.	surface primitives		classification criterions				
	Name	canonical expression	ϕ	Φ	Δ	k-sign	K-sign
1	One real plane	$ax + by + cz + d = 0$	1	1			
2	Ellipsoid	$x^2/a + y^2/b + z^2/c = 1$	3	4	< 0		
3	elliptic cylinder	$x^2/a + y^2/b = 1$	2	3		same	oppo.
4	hyperbolic cylinder	$x^2/a - y^2/b = 1$	2	3		oppo.	
5	parabolic cylinder	$x^2 + 2y = 0$	1	3			
6	quadric cone	$x^2/a + y^2/b - z^2/c = 0$	3	3		oppo.	
7	hyperboloid of one sheet	$x^2/a + y^2/b - z^2/c = 1$	3	4	> 0	oppo.	
8	hyperboloid of two sheets	$x^2/a + y^2/b - z^2/c = -1$	3	4	< 0	oppo.	
9	hyperbolic paraboloid	$x^2/a - y^2/b + 2z = 0$	2	4	> 0	oppo.	
10	elliptic paraboloid	$x^2/a - y^2/b + 2z = 0$	2	4	< 0	same	
11	intersecting planes	$x^2/a - y^2/b = 0$	2	2		oppo.	
12	parallel planes	$x^2 = 1$	1	2			oppo.

(*** "oppo." means "opposite"; $a, b, c \neq 0$.**)

MODEL MERGING.

After model estimation, the PCDs from some groups may be from the same surface. Model merging is a procedure of merging the points from same surface primitives then re-estimating the quadric model. Model merging is performed in iterative manner. A binary adjacent matrix W is first generated. If i-th and j-th groups are adjacent then w_{ij}=1; otherwise w_{ij}=0. W is a symmetric matrix with the diagonal elements being one. Check all the adjacent pairs of groups in W. If the classified shapes of two adjacent groups are the same, then the Mahalanobis norm between these two models is computed:

$$dist(\theta_i, \theta_j) = \|\theta_i - \theta_j\|_{\Sigma} = (\theta_i - \theta_j) \cdot \Sigma \cdot (\theta_i - \theta_j)^T \qquad (12)$$

where $\theta_i, \theta_j \in \mathbb{R}^{10 \times 1}$ are estimated parameter vector of two adjacent groups; Σ is a diagonal matrix with the diagonal elements as [0.15, 0.15, 0.15, 0.1, 0.1, 0.1, 0.05, 0.05, 0.05, 0.1]. Note that this norm assigned different weights to different model parameters in the model comparison. If $dist(\theta_i, \theta_j)$ is smaller than a threshold, the points from these two groups are merged. Meanwhile, the adjacent matrix is updated accordingly and a quadric model is re-estimated using the new group. The model merging continues until no more adjacent groups need to be merged. Finally, the inlier sets for each merged model are determined from the whole PCD.

EXPERIMENT AND EVALUATION

This section applies the algorithm to a bridge whose surfaces include a non-planar element. The three evaluation metrics used to evaluate the proposed algorithm are the *algebraic fitting error,* the *point-to-surface (PS) distance*, and the *detection percentage*. The fitting error is $E_{Alg} = \theta \cdot [x^2, y^2, z^2, yz, zx, xy, x, y, z, 1]^T$, where θ is the estimated model parameter vector. The PS distance is the mean distance from all inlier points to the estimated surface. Computing the PS distance requires finding the closest point $[x^*, y^*, z^*]^T$ on the surface. Assume the surface model is defined implicitly by $f(x, y, z) = 0$, then the closest point $[x^*, y^*, z^*]^T$ on the surface to a data point $[x_i, y_i, z_i]^T$ is the solution to the optimization problem:

$$\min_{x^*, y^*, z^*} \quad dist = (x^* - x_i)^2 + (y^* - y_i)^2 + (z^* - z_i)^2, \qquad (13)$$
$$s.t. \quad f(x_i, y_i, z_i) = 0$$

with Lagrange Multiplier this can be solve by:

$$\frac{\partial F}{\partial x^*} = 0 \; ; \frac{\partial F}{\partial y^*} = 0 \; ; \frac{\partial F}{\partial z^*} = 0 \; ; \frac{\partial F}{\partial \lambda^*} = 0 \qquad (14)$$

where $\quad F = (x^* - x_i)^2 + (y^* - y_i)^2 + (z^* - z_i)^2 + \lambda \cdot f(x, y\_i, z) \qquad (15)$ Eq. (15) leads to a sextic equation with six solutions. The correct solution is the one with the smallest real value of PS distance.

The detection percentage measures how completely the algorithm is able to extract the models. For example, if there are N models in the PCD and the algorithm detects M models accurately, then the detection coverage is $100 \, M/N$. The ideal result is 100%.

The proposed algorithm is tested using a PCD of a real-world bridge, shown in **Figure 2**. The PCD is of 407503 points and reconstructed from image sequences using CMVS (Furukawa, 2010). In this PCD, there are two planes from the walls on both sides and one cylinder between the two planes. As part of the first step, the PCD

is down-sampled to 7756 points, shown in **Figure 3 (Left).** The segmentation step uses a coding length threshold as 15 and over-segments the PCD into 8 groups. In model estimation, $\gamma = 0.9$, $\sigma = 0.5$. For the classification step, the normalized truncation threshold is 0.005. The classification result before model merging is: three "one real plane" models and four "parabolic cylinder" models and one "hyperbolic paraboloid" model. Then the model merging is performed and finally 8 groups are merged into 3 groups: two "one real plane" models and one "parabolic cylinder" model, as shown in **Figure 3 (Right)**. The experiment is performed on a PC with a 2.80GHz Intel Core i5 CPU and 8GB RAM, using C++ and MATLAB. The down-sampling step is in C++ and the other steps are in MATLAB. In total, the experiment took 452.13 seconds (the segmentation took 201.90 seconds). For each model, the algorithm returns the estimated model parameters. The quantitative evaluation is listed in Table 2 result using the evaluation metrics discussed in Section 4.1. Moreover, the detection percentage is 100%.

Figure 2. The PCD of a real bridge used in the experiment
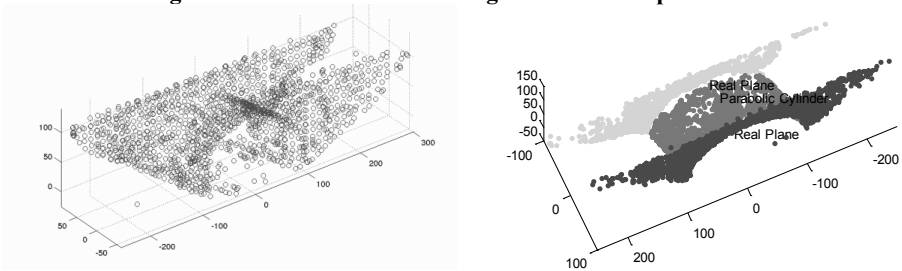
Figure 3. Results on the bridge PCD. Left: the down-sampled PCD; Right: final detection and classification results

Table 2. Mean PS distance for experimental results on the bridge PCD.

Surface Detected	Real Plane #1	Parabolic Cylinder	Real Plane #2
Algebraic fitting error	0.0235	5.3943e-04	0.0242
Mean of PS distance (cm)	1.2344	1.3248	1.9129

CONCLUSION

A novel algorithm for detecting, fitting and classifying quadric surface primitives for infrastructure PCDs is presented in this paper. The algorithm consists of several steps, including fast segmentation, pose recovery, quadric model estimation, quadric surface classification, and model merging. Among these steps, the pose recovery, model estimation and model merging are new algorithms proposed while the other two are modified from existing state-of-the-art methods. Evaluation of the

algorithm used three quantitative evaluation metrics, i.e. algebraic fitting error, mean point-to-surface (PS) distance and detection percentage. The algorithm was tested on a PCD of a real bridge, generated by CMVS techniques, and shown to accurately recover the underlying surfaces.

REFERENCES

Zhang, G., Karasev, P., Brilakis, I., Vela, P.A. (2012). Sparsity-Inducing Optimization Algorithm for the Extraction of Planar Structures in Noisy Point-Cloud Data. *Proc. IWCCE*, July.

Sundar, H., Silver, D., Gagvani, N., Dickinson, S. (2003), Skeleton based shape matching and retrieval. *Proc. Shape Modeling Applications Conf.,* June.

Tung, T., Schmitt, F. (2004). Augmented Reeb graphs for content-based retrieval of 3D mesh models. *Proc. Shape Modeling Applications Conf.,* June.

Kuo, C. T., Cheng, S. C. (2007). 3D model retrieval using principal plane analysis and dynamic programming. *Pattern Recognition*, *40*(2): 742-755.

Tangelder, J. W., Veltkamp, R. C. (2008). A survey of content based 3D shape retrieval methods. *Multimedia Tools and Applications*, *39*(3): 441-471.

Drost, B., Ulrich, M., Navab, N., Ilic, S. (2010). Model globally, match locally: Efficient and robust 3D object recognition. *Proc. CVPR*, June.

Johnson, A. E., Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3D scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(5): 433-449.

Chen, H., Bhanu, B. (2007). 3D free-form object recognition in range images using local surface patches. *Pattern Recognition Letters*, 28(10): 1252-1262.

Rusu, R. B., Blodow, N., Beetz, M. (2009). Fast point feature histograms (FPFH) for 3D registration. *Proc. ICRA*, May.

Mian, A. S., Bennamoun, M., Owens, R. (2006). Three-dimensional model-based object recognition and segmentation in cluttered scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(10): 1584-1601.

Rabbani, T., Van Den Heuvel, F. (2005). Efficient Hough transform for automatic detection of cylinders in point clouds. *ISPRS WG III/3, III/4*, (3): 60-65.

Zaharia, T., Preteux, F. J. (2001). Hough transform-based 3D mesh retrieval. *Proc. International Symposium on Optical Science and Technology*, November.

Schnabel, R., Wahl, R., Klein, R. (2007). Efficient RANSAC for Point-Cloud Shape Detection. *Computer Graphics Forum*, 26(2): 214-226, June

Ma, Y., Derksen, H., Hong, W., Wright, J. (2007). Segmentation of multivariate mixed data via lossy data coding and compression. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(9): 1546-1562.

Torr, P. H., Zisserman, A. (2000). MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1), 138-156.

Choi, S., Kim, T., Yu, W. (2009). Performance evaluation of RANSAC family. In *Proc. British Machine Vision Conference*, Sept.

Furukawa, Y., Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8): 1362-1376.

Formalized Approach for Accurate Geometry Capture through Laser Scanning

Anu R. Pradhan, A.M.ASCE[1], Franklin L. Moon, A.M.ASCE[2]

[1]Assistant Professor, Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA 19104; PH (215) 571-3540; FAX (215) 895-1363; email: pradhan@drexel.edu
[2]Associate Professor, Department of Civil, Architectural and Environmental Engineering, Drexel University, Philadelphia, PA 19104; PH (215) 895-6099; FAX (215) 895-1363; email: flm72@drexel.edu

ABSTRACT

The main objective of the presented research is to formalize a novel iterative feedback mechanism that facilitates reliable and accurate geometric capture of civil infrastructure systems (e.g., bridges) through laser scanning to support predictive modeling (e.g., finite element modeling). Current condition assessment (of a bridge) based on visual inspection is subjective, and often fails to obtain reliable and accurate geometry data which is essential for finite element modeling. The laser scanner, which is a non-contact geometry capture technology, is one of the promising technologies that can capture more detailed (2mm spacing) and accurate (+/- 5mm) geometric data compared to existing surveying and measurement techniques, such as total stations and Global Positioning Systems. The presented work is focused on investigating and formalizing how laser scanning technology can optimally be leveraged and integrated to provide a sufficiently accurate documentation of bridge geometry, and then how such information may be transformed into a reliable finite element model for predictive modeling.

Keywords: Laser Scanner, Simulation, Structural Health Monitoring

1. INTRODUCTION

The main objective of the presented research is to formalize a novel simulation-based framework that facilitates reliable geometric capture of civil infrastructures through a 3D geometry capture device (e.g., laser scanner). Embedding a simulation model within the control-loop of a 3D geometry capture device is a novel proposition, and has the potential to greatly improve both the efficiency and robustness of current geometry capture approaches. Achieving this vision will require the confluence of multiple interdisciplinary domains, such as probabilistic error modeling, finite element modeling and data mining.

Over the last few decades there has been significant interest in the acquisition and use of quantitative data to assess the current performance and forecast the future performance of civil infrastructure. This field, broadly termed Structural Health Monitoring (SHM), has focused primarily on the use of kinematic responses, such as strain, displacement, and acceleration. While such responses are of great importance, they do not sufficiently characterize the performance of a structure, nor do they

provide sufficient information to forecast all meaningful performances. The need to broaden SHM to consider more diverse data streams is a primary motivation for the proposed research, which aims to establish an efficient approach to accurately capture geometric information of civil infrastructures.

Current condition assessments of our critical civil infrastructures are based exclusively on visual inspection. In the case of bridge, inspectors assign subjective condition ratings based on visual appearance alone and these ratings are used to prioritize interventions. There is ample evidence that the condition ratings based on such visual inspection, on average, are found to be highly variable (Phares et al. 2004). Further, in many cases the desired performance metric (e.g., structural capacity) is often poorly correlated with visual appearance since estimating such structural capacity requires estimates of as-is geometry, especially noting any permanent deformations or distortions in the critical elements that may affect their remaining safety and stability. Although a few experts may be able to do such estimation reliably through visual inspection, in general such estimates require the capture of accurate geometry and the use of simulation modeling approaches, such as finite element (FE) modeling, to inform safety limit state assessments. Even if an inspector may be an experienced structural engineer, given the limitations of human experience, sight and perception, most inspectors cannot be expected to be able to notice critical geometric anomalies and other signs of distress reliably. Further, even if a bent gusset plate may be noticed during an inspection (as in the case of the I-35W bridge that collapsed in Minnesota), connecting this to the likely failure modes and the likelihood of in-service structural failure is challenging with today's engineering practice.

The laser scanner, which is a non-contact 3D geometry capture device, can acquire geometric information with the spatial resolution (2mm spacing) and accuracy (+/- 4mm) needed to generate representative simulation models. The advent of such geometry-capture devices provides a unique opportunity to remotely capture three-dimensional (3D) as-is geometry of structures, and thus complement SHM systems based on both model-based and data-driven interpretation frameworks. From an infrastructure inspection perspective, what is important is investigating how this technology can optimally be leveraged and integrated to provide a sufficiently accurate documentation of bridge geometry, and then how such information may be transformed into a reliable FE model. Once a reliable FE model that represents the current as-is geometry becomes available, we can expect the next generation of structural analysis software to perform a large number of scenario analysis to estimate the reliability of a bridge at routine intervals or following an extreme event (e.g., earthquake or hurricane). To realize this potential, fundamental research challenges associated with the acquisition and interpretation of geometric information need to be addressed.

The motivation behind the presented research follows from the vision above. This research is aimed at exploring how to fuse local and global geometric data captured through the use of a 3D imaging system (e.g., laser scanner) with simulation modeling to estimate the key performance metrics (e.g., structural capacity) given the as-is geometry. Although a laser scanner is a promising technology in terms of acquiring accurate geometry data, the accuracy of (geometric) data depends on the

distance and incidence angle between the scanned object and the scanner as well as the shape and material properties of the scanned object. For instance, increasing the distance and incidence angle between the scanned object and the scanner introduces substantial measurement errors in the data (Hebert et al. 1991). At the same time, when one considers the predictions of certain performances, it is important to recognize that the geometric data of certain portions of a structure are more critical than the geometric data of other portions of a structure. During data collection, it is often difficult for field personnel to reliably identify the critical portions of the structure that may require re-positioning of the scanner and perform multiple scans to capture geometric data accurately. Thus, to streamline the data collection process and transform (laser) scanned data into an accurate FE model, a comprehensive framework that considers different error sources and geometric measurements at critical portions of a structure is needed. Such a framework does not currently exist, and formalizing such a framework is the main motivation of this proposed work. The presented approach, not limited to bridge condition assessment, has broader applications in terms of assessing other civil constructed systems (e.g., dams and buildings), preservation and restoration of historical structures, construction quality control, and inspection of cultural artifacts and industrial design.

2. EXISTING RESEARCH STUDIES

Research studies related to data acquisition have focused on studying the characteristics of scanners (Cheok et al. 2003; Lichti et al. 2006; Tang et al. 2009) and optimizing the positions of scanners to reduce the number of scans required (Latimer et al. 2004). Similarly, research studies on data processing are focused on the issues related to registration of multiple laser scans and data noise reduction techniques (Huber et al. 2003). Under data interpretation, the research emphasis has been on the areas of geometric modeling and object recognition (Zhang et al. 1997; Teizer et al. 2007; Bosche et al. 2008).

Although the current research has produced important advancements, all studies have assumed a linear process without any feedback mechanism that allows the data interpretation activities to inform the data acquisition activities (Figure 1). In other words, during data acquisition, there is no information on whether the collected data is useful to support a given task due to high degree of uncertainty in measurements due to errors, missing information, etc. Although promising sensing technologies, such as laser scanners, have created unique opportunities to assess civil infrastructure systems, the lack of a feedback mechanism is currently limiting the effectiveness and efficiency with which such technologies may be employed (Bosche et al. 2008).
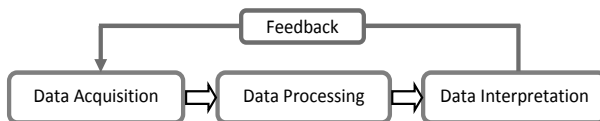
Figure 1 Envisioned Feedback System

The concept of a feedback mechanism is extensively used in control theory to deal with the behavior of dynamical systems (Nise 2008). The basic concept is that when one or more output variables of a system need to follow a certain desired output, a controller manipulates inputs to the system in order to produce the desired effect. Biological systems, such as organisms and ecosystems also contain some type of feedback mechanism that enhances system functions.

3. RESEARCH VISION
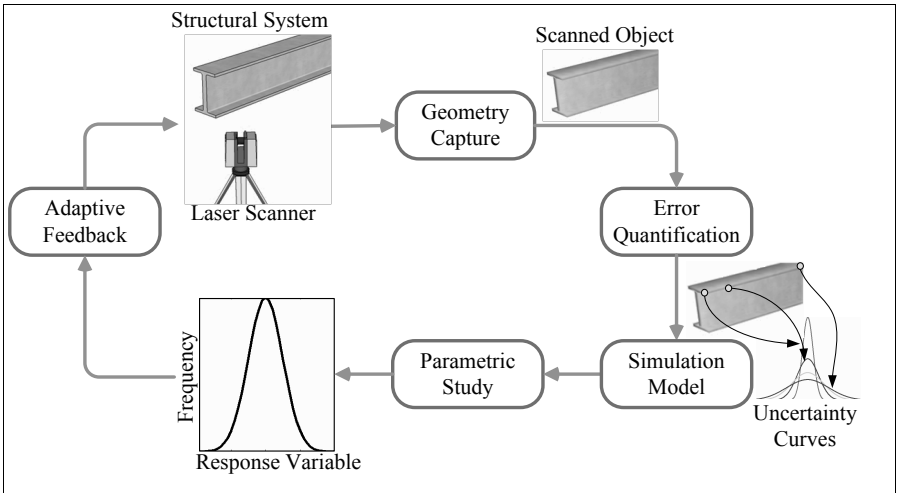
Figure 2 Research Vision

We envision an adaptive feedback mechanism (shown in Figure 2) to enable robust and intelligent geometry capture. Using this approach an inspector would capture comprehensive, but coarse geometric information of a structural system using a laser scanner. The scanned data would then be analyzed to quantify the measurement errors (via a probabilistic sensor measurement model) at different sections of the structure since measurement errors depend on parameters that would vary spatially across the structure (e.g., distance and incidence angle). The simulation model would then perform predictive modeling of the structure based on the geometric measurements and measurement errors of structural components. The simulation model would then develop a histogram of the desired performance/response metric (e.g., deflection, demand/capacity rations, stress concentrations, etc.) that reflects the level of uncertainty within the geometric data. If the response metric and its associated uncertainty are deemed acceptable, then no additional scans would be required.

However, if the uncertainty associated with the response metric is unacceptable, the adaptive feedback module would analyze the acquired data to identify the geometric regions (of the structure) that were contributing most significantly to the variability of the response metric. This information would then be provided to the user in the form of a set of recommended scanner locations to most

effectively reduce the uncertainty associated with the predicted response metric. Thus, the feedback mechanism would enable the acquisition of geometric data with a resolution and accuracy that reflected (1) the importance of the region being captured, and (2) the acceptable level of uncertainty associated with the desired response prediction.

4. PRELIMINARY FINDINGS

The research team conducted a preliminary study to assess the as-is geometry of a bridge located at Wayne, NJ. The research team was interested in creating a 3D as-is model of the bridge, and converting such a model into a finite element model. The research team scanned the bridge using the Leica^{TM} ScanStation C10 laser scanner under the supervision of professional surveyor from Pennoni Associates Inc. Figure 3 shows the as-built model of the bridge, which was created by fitting different geometric shapes (e.g., I-section, cylinders, planes) to the point cloud data obtained from the laser scanner. The commercially available Leica^{TM} Cyclone software was used to create the 3D as-is model. The research team measured the geometric dimensions of different bridge components (e.g., piers, pier caps, girders), and compared the dimensions with actual design drawings.
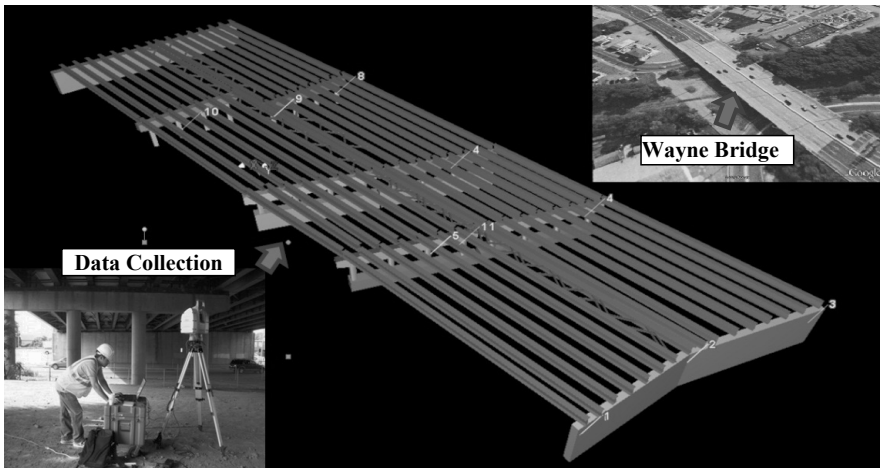
Figure 3 3D As-Is Model of Wayne Bridge, NJ

The comparison results showed that the difference between the actual measurements and measurements obtained from the laser scanner varied from 8 millimeters (mm) to 28mm. The difference was found to be minimum (i.e., 8mm) for cylindrical pier columns and maximum (28mm) for some I-beam girders. Although the advertised measurement error of the laser scanner (based on manufacturer's specifications) is about 4-5 mm, the reported error seems to be reasonable for only cylindrical pier columns. Existing research studies have identified multiple factors (e.g., scanning distance, incidence angle, shape and material of scanned object) that can influence the measurement accuracy of scanned data. In the abovementioned

comparisons, <u>it was not clear which factors contributed most to the measurement errors</u>. In such situation, an approach that can guide a user (e.g., data collector) to minimize possible measurement errors is desired, and currently such an approach is not available. In addition to minimizing measurement errors, <u>another important element that is currently overlooked is the importance of geometric data at certain portions of a structure</u>. The geometric data of certain portions of a structure are more critical than geometric data of other portions of a structure when considering certain performance/response metrics. Thus, the uncertainty in geometric measurement at those critical portions of a structure can substantially affect the accuracy of specific response metrics from simulation (e.g., finite element modeling). To demonstrate this concept, we present a simple example using a cantilever beam as shown in Figure 4.
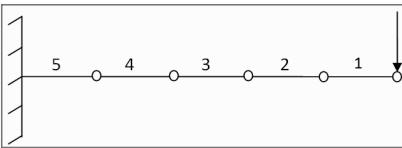
Figure 4 Cantilever Beam under Point Load

The deflection of the beam at the free end under 1 kip of load was selected as the desired performance/response metric. For the modeling, the beam was divided into five different sections of equal length (L) along its longitudinal axis. Each section was given a numeric identification number starting from 1 (at the free end) to 5 (near the fixed support) as shown in Figure 4. For the simulation, the sectional length was taken as 25 in., Young's modulus (E) was taken as 29,000 ksi, and the mean height and breath of the beam were taken as 6 in. and 1 in., respectively. Using the mean values for the cross-sectional dimensions the deflection under a 1 kip point load is 1.2472 in.

To illustrate the proposed concept, the cross-sectional dimensions of the beam were assumed to be captured by a laser scanner. During the simulation, measurement samples were generated by varying the error model for the height and breadth measurements for each section of the beam. For simplicity, the errors were assumed to be normally distributed with a mean equal to the nominal dimensions, and the first standard deviation varied to examine the spatial influence of errors. At first, a suite of 1000 measurement samples were generated by assigning a standard deviation of 10% to all error models. Subsequently, additional measurement samples were generated by reducing the standard deviation associated with error models for individual beam sections (either height or breadth) as if the scanner was located so as to more accurately capture certain geometry. The generated measurement samples were then fed into a finite element model to compute the histogram of the desired performance/response metric (i.e., deflection in this example).

Figures 5 (left) and (right) show the resulting displacement predictions when the measurement errors for each section (breadth and height measurements, respectively) are reduced one at a time (a reduction in the width of the distribution signifies a reduction in uncertainty). It is evident that reducing measurement errors at Section 5 of the beam (blue dashed curve in Figure 5) greatly reduces the uncertainty in response prediction compared to the other sections of the beam. Of course in this example it is clear that this section experiences the largest moment (and undergoes the largest curvature) and thus contributes most significantly to the tip displacement.

However, in general it is not always immediately apparent, which regions most significantly influence the desired prediction. In addition, Figure 4 also illustrates that the height measurement is also far more influential than the width measurement (given the prescribed loading). Similarly, in this case such a result may be intuitively obvious (since the moment of inertia is proportional to the product of the height cubed and the breadth) but in general this is not the case. The simplicity of this example serves as a clear illustration that not all geometric data is of equal importance. More specifically, two general conclusions can be drawn:

(1) Geometry measurements at certain portions of a structure (e.g., Section 5) may be far more critical than geometry measurements of other portions (e.g., section 1) of a structure.

(2) A certain geometry measurement (e.g., height measurement) of a specific portion of a structure may be far more critical than a different geometry measurement (e.g., breadth measurement) of the same portion of a structure.
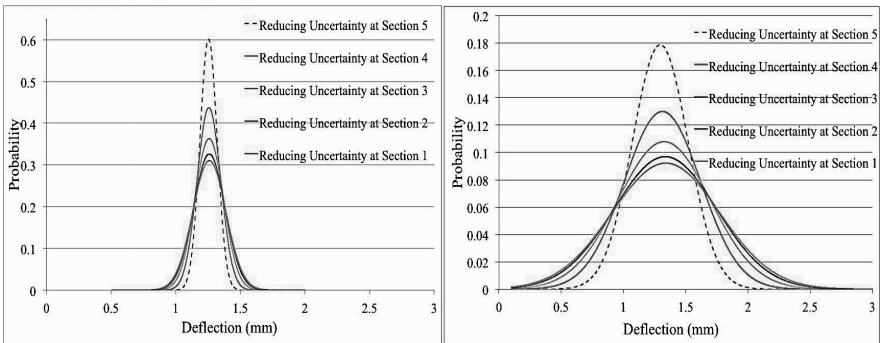
Figure 5 Plots showing Uncertainty in Simulations Due to Measurement Errors at Different Sections

How these two observations transfer to other structures depends greatly on the structural system being considered (load paths, force-resisting mechanics, etc.) and the desired response/performance metric. Although it is difficult even for structural engineering experts to identify such interactions based on intuition, it is a straightforward task for simulation models. It is this synergy that underpins the envisioned novel feedback mechanism proposed herein.

5. CONCLUSIONS

The presented research highlighted the need for a novel iterative feedback mechanism that facilitates reliable and accurate geometric capture of civil infrastructure systems (e.g., bridges) through laser scanning to support predictive modeling. The need for such a framework is based on the preliminary investigations conducted by the research team. The presented work is focused on investigating and formalizing how laser scanning technology can optimally be leveraged and integrated to provide a sufficiently accurate documentation of bridge geometry, and then how

such information may be transformed into a reliable finite element model for predictive modeling.

6. REFERENCES

Bosche, F., and Haas, C. T. (2008). "Automated retrieval of 3D CAD model objects in construction range images." *Automation in Construction*, 17(4), 499-512.

Cheok, G., Stone, W., and Witzgall, C. (2003). "Some Issues Relating to Performance Evaluation of LADARs." *Proc., Proceedings of PerMIST-03*.

FHWA (2002). "Reliability of Visual Inspection for Highway Bridges." Federal Highway Administration

Gong, J., and Caldas, C. H. (2008). "Data processing for real-time construction site spatial modeling." *Automation in Construction*, 17(5), 526-535.

Hebert, M., and Krotkov, E. (1991). "3-D measurements from imaging laser radars: how good are they?" *Proc., Proceedings of the IEEE/RSJ International Workshop on Intelligent Robots and Systems IROS '91*, IEEE, 359-364.

Huber, D. F., and Hebert, M. (2003). "Fully automatic registration of multiple 3D data sets." *Image and Vision Computing*, 21(7), 637-650.

Laefer, D. F., and Pradhan, A. R. (2006). "Evacuation Route Selection Based on Tree-Based Hazards Using Light Detection and Ranging and GIS." *Journal of Transportation Engineering*, 132(4), 312-320.

Latimer, E., Saxena, R., Lyons, C., Michaux-Smith, L., and Thayer, S. (2004). "Sensor space planning with applications to construction environments." Institute of Electrical and Electronics Engineers Inc., Piscataway, United States, 4454-4460.

Lichti, D., Pfeifer, N., and Maas, H.-G. (2008). "ISPRS Journal of Photogrammetry and Remote Sensing theme issue "Terrestrial Laser Scanning"." *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(1), 1-3.

Nise, N. S. (2008). *Control systems engineering*, Wiley, Hoboken, NJ.

Phares, B. M., Washer, G. A., Rolander, D. D., Graybeal, B. A., and Moore, M. (2004). "Routine Highway Bridge Inspection Condition Documentation Accuracy and Reliability." *Journal of Bridge Engineering*, 9(4), 403-413.

Tang, P., Akinci, B., and Huber, D. (2009). "Quantification of edge loss of laser scanned data at spatial discontinuities." *Automation in Construction*, 18(8), 1070-1083.

Teizer, J., Caldas, C. H., and Haas, C. T. (2007). "Real-Time Three-Dimensional Occupancy Grid Modeling for the Detection and Tracking of Construction Resources." *Journal of Construction Engineering and Management*, 133(11), 880-888.

Zhang, D., and Hebert, M. (1997). "Multi-scale classification of 3-D objects." *Proc., Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 864-869.

Potentials of RGB-D Cameras in As-built Indoor Environments Modeling

Zhenhua Zhu[1] and Sara Donia[2]

[1]Department of Building, Civil, and Environment Engineering, Concordia University, Montreal, Canada, H3G 1M8; PH (514) 848-2424 ext. 5948; FAX (514) 848-7965; email: zhenhua.zhu@concordia.ca

[2]Department of Building, Civil, and Environment Engineering, Concordia University, Montreal, Canada, H3G 1M8; email: sa_donia@hotmail.com

ABSTRACT

3D as-built models of building indoor environments could be used to facilitate multiple building assessment and management tasks, such as post-disaster safety evaluation, renovation/retrofit planning, and maintenance scheduling. However, modeling building indoor environments is a challenging task. It is even more difficult than modeling building facades due to the issues, such as limited lighting conditions and prevalence of texture-poor walls, floors, and ceilings. This paper investigates the potentials of RGB-D cameras in modeling building indoor environments. Three pilot studies have been performed to evaluate 1) the accuracy of the sensing data provided by an RGB-D camera and 2) the automation that can be achieved for the registration of building indoor scenes and the recognition of building elements with the sensing data. The studies show that the camera can provide a stream of mid-accurate sensing data in real time. Also, a high degree of automation can be achieved through the fusion of spatial and visual data from the camera, when modeling the as-built conditions in building indoor environments.

KEYWORDS: RGB-D Camera; As-built Modeling; Indoor Environments

INTRODUCTION

3D as-built models of building indoor environments record the existing conditions in buildings, including the actual details of architectural and structural elements. Therefore, the models are useful to facilitate multiple building assessment and management tasks, including but not limited to post-disaster safety evaluation, renovation/retrofit planning, and maintenance scheduling. For example, the use of 3D as-built models could significantly facilitate the coordination of the mechanical, electrical, and plumbing (MEP) design, when renovating and retrofitting existing old buildings. This facilitation was expected to reduce almost 60% of the MEP field-to-finish workflow (ClearEdge3D, 2012).

Although the 3D as-built models of building indoor environments are useful, the process for automatically creating these models is not easy. It is even more challenging than modeling building facades. Building indoor environments are commonly prevalent with texture-poor walls, floors, and ceilings. Considering these texture-poor elements plus the limited lighting conditions, Furukawa et al. (2009) pointed out that most automated modeling methods based on images or videos would fail unless the scene-specific constraints were pre-created. Also, the existence of building interior partitions makes the scans of building indoor environments with the sensing devices like laser scanners more time-consuming and labor intensive, since the laser scanners have to be set up, moved, and re-set up tens or hundreds of times.

The emergence of RGB-D camera provides a promise to create as-built models of building indoor environments in an automatic, fast and efficient way. One example is the Microsoft® Kinect. The camera is expected to be able to capture a real-time stream of color and depth images, which can be converted to a set of 3D colored point clouds. Also, the lightweight and portable nature of the camera makes it convenient to use, when modeling the building indoor environments with multiple partitions. So far, several research studies have been proposed, but most of them just focused on how to reconstruct building indoor scenes in 3D with the sensing data captured by the Kinect camera (Henry et al. 2012; Izadi et al. 2011).

The generation of as-built models is more than the 3D scene reconstruction. It also includes the recognition and organization of building elements in the scenes. In addition, the accuracy and resolution of the sensing data directly affect the final as-built modeling results. The objective of this paper is to investigate the potentials of the Kinect camera in these issues. Specifically, the focus in this paper is placed on evaluating: 1) the accuracy of the sensing data that the camera can provide, and 2) the automation that can be achieved for the registration of building indoor scenes and the recognition of building elements with the sensing data from the camera.

BACKGROUND

The benefits of as-built building models have been well acknowledged by researchers and professionals in the architecture, engineering, and construction industry. Meanwhile, the process of creating such models has been identified as labor-intensive and time-consuming (Tang et al. 2010). In order to facilitate the generation of as-built building models, many automated methods have been developed. These methods can be classified into three main categories based on the sensing data they use. The methods in the first category are built upon the 3D points captured directly by the devices like terrestrial laser scanners, which could collect millions of 3D points with one scan in minutes. The methods in the second category rely on the digital images or videos taken by digital cameras or camcorders. The digital cameras or camcorders are easy to use and portable, but the images or videos are 2D. Therefore, the 3D information has to be estimated from multiple images or video frames shot from different directions. Recently, the methods using RGB-D cameras have been proposed. The RGB-D cameras can capture color and depth images simultaneously and in real time. One example is the Kinect developed by Microsoft®. Below are the details about the methods in three categories.

Point cloud based modeling

The devices like terrestrial laser scanners can capture the detailed as-built conditions of building elements in the form of a dense set of 3D points. Based on these points, several modeling methods have been developed. For example, Okorn et al. (2010) proposed an idea of modeling building floor plans by projecting the collected 3D points onto a vertical (z-) axis and a ground (x-y) plane. The projection results could indicate which points can be grouped to form floor plans (Okorn et al. 2010). Xiong and Huber (2010) employed the conditional random fields to model building elements from 3D points, but their work is limited to modeling the elements with planar patches, such as walls. The openings on the walls could be further located by comparing point density or using a support vector machine (SVM) classifier (Adan and Huber, 2011).

Although the detailed information can be captured with a laser scanner, it is heavy and not portable. Typically, at least two crews are needed in the sensing data collection process (Foltz, 2000). This non-portable nature makes the laser scanner not convenient to use when modeling building indoor environments. Adan et al. (2011) mentioned that a scanner had to be set up at 225 different locations in order to scan 40 rooms (i.e. approximately 5.6 locations per room). Also, the 3D points collected by the laser scanner only record the spatial information of indoor scenes. Other building information was lost. Take material information as an example. It is difficult to differentiate a concrete column or a wooden column just from their 3D points, if both have the same shape and size.

Image or Video based modeling

The main idea of image or video based modeling is to retrieve 3D information from multiple 2D images or video frames. So far, several critical techniques have been created in computer vision to make this idea possible and practical, including the structure-from-motion (Snavely et al. 2006), multi-view stereo matching (Furukawa and Ponce, 2010), and simultaneous location and mapping (Durrant-Whyte and Bailey. 2006). Built upon these techniques, Geiger et al. (2011) established a framework which could perform the 3D reconstruction of built environments in real time. Pollefeys et al. (2008) modeled urban built environments using the videos from a moving vehicle.

Compared with laser scanners, digital cameras or camcorders are easy to use and portable. Therefore, the cameras are convenient to collect as-built conditions in building indoor environments especially with multiple interior partitions. However, most image or video based modeling methods heavily relied on the visual features extraction from the images or video frames. The building indoor environments are commonly prevalent with texture-poor walls, floors, and ceilings. This may fail the procedures of these modeling methods, unless the scene-specific constraints for the environments are pre-created manually (Furukawa et al. 2009).

RGB-D based Modeling

The modeling methods in the aforementioned two categories have their own limitations when modeling the as-built conditions in the indoor environments. Most of these limitations are inherent to the sensing systems they used. The emergence of

RGB-D cameras provides another way to create as-built models. The RGB-D cameras are novel sensing systems, which can capture mid-resolution color and depth images at high capture rates. For example, the resolution for the Microsoft® Kinect camera could be 640x480, which equals to the collection of 307,200 3D points per frame. Considering the 30 frames-per-second (FPS) sensing rate, almost 1 million 3D points can be captured in one second.

Based on the color and depth images collected by the RGB-D cameras, several methods have been developed. Henry et al. (2010) investigated how such cameras can be used for building a dense 3D map of building indoor environments. In their method, the registration of 3D points was performed by matching the sparse key points in consecutive color images (Henry et al. 2012). Fioraio and Konolige (2011) used the bundle adjustment to align 3D points directly without exploiting color images. More impressively, Izadi et al. (2011) developed a novel framework, KinectFusion, which enables a user to hold and move a standard Microsoft Kinect camera to rapidly and automatically create a detailed 3D reconstruction for indoor scenes. However, their work is currently limited to a relatively small and bounded area fixed in the world (Roth and Vona, 2012; Whelan et al .2012).

Although the preliminary results from the recent research work related to RGB-D based modeling are promising, most current studies just focused on the 3D reconstruction of building indoor scenes. The full potentials of the RGB-D cameras for modeling as-built conditions in the building indoor environments are not well investigated. Several problems are not well answered. For example, how accurate could the 3D reconstruction of building indoor scenes points reach? What are the benefits and limitations of using color information in the registration of 3D scenes? What kind of building elements could be recognized with the sensing data provided by the RGB-D cameras?

OBJECTIVE AND SCOPE

The objective of this paper is to investigate the potentials of the Microsoft® Kinect camera, one of common RGB-D cameras in the market, for modeling as-built conditions in building indoor environments. The focus is placed on investigating: 1) the accuracy of the sensing data provided by the Kinect; 2) the difference of the 3D scenes registration with/without color images; and 3) the possibility of using color information to recognize building elements in 3D. Three pilot studies have been performed. The 3D points from the Kinect camera are compared with the results from a Total Station to indicate the sensing accuracy of the Kinect camera. The effectiveness of the scene registration with and without exploiting visual features in color images are compared. In the last study, concrete column surfaces are recognized from the 3D point clouds using the visual features contained in color images.

PILOT STUDIES

Study I: Determination of Kinect Sensing Accuracy
The Kinect sensing accuracy is determined by comparing the 3D points from the Kinect with the ones from a Total Station instead of a laser scanner (Khoshelham

and Elberink, 2012), considering the difference in measurements between a total station and a laser scan could be 0.2 - 2 cm (Haddad and Ishakat, 2007). Specifically, control points in a building indoor scene are first identified. According to the scan of the scene with the Kinect, the 3D coordinates of the control points are estimated. Also, the point coordinates are obtained with a Total Station. Consider the Kinect and Total Station may have different coordinate systems, the comparison is not directly made between the point coordinates measured by the Kinect and Total Station. Instead, the distance value between any two control points is calculated, and the sensing accuracy of the Kinect is determined by comparing the distance values.

Figure 1 illustrates the experiments performed in this study to determine the sensing accuracy of the Kinect. Table 1 shows two examples of the distance values measured by the Kinect and Total Station separately. According to the results form 70 experiments, the average absolute and percentage errors are 3.6 cm and 2.898%.

(a) Survey with a Total Station

(b) Survey with a Kinect camera

(c) Control points identified in the image
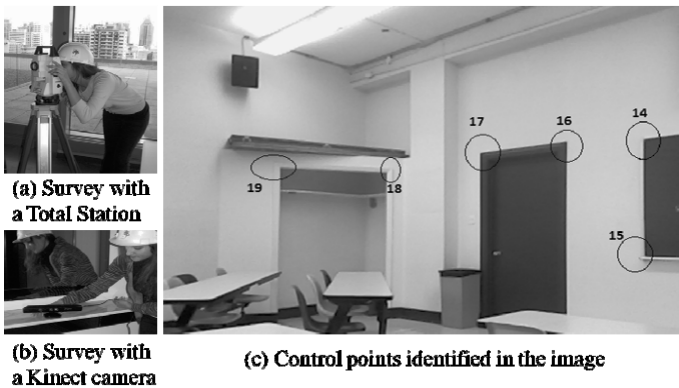
Figure 1. Experiments for Determining the Kinect Sensing Accuracy.

Table 1. Examples of Sensing Data Collected.

	No. 34	No. 35
Distance – Total station	0.843	4.123
Distance - Kinect	0.818	4.166
Absolute error (meter)	0.025	0.043
Percentage error	2.97%	1.04%

Study II: 3D Scene Registration

Scene registration refers to the merge and alignment of 3D point clouds from multiple scans into one single point cloud under a pre-defined coordinate system. It is always necessary, because one scan typically cannot capture all environmental details. This is especially true for RGB-D cameras, since they have limited image resolutions.

One critical step in the 3D scenes registration is to estimate the pair-wise alignment matrix between two consecutive point cloud scans. Typically, the matrix is

estimated by finding and matching the 3D points that are common in both scans. Considering the Kinect camera can simultaneously provide color and depth images, two categories of methods are developed to match the common 3D points in two point clouds. The methods in the first category only use the depth information. Figure 2a illustrates one example of such methods. In this example, the common 3D points between two point clouds are found by matching the 3D spatial features at the key points with a fast point feature histograms (FPFH) descriptor (Rush, 2010). The methods in the second one rely on both color and depth information. In these methods, the 2D visual features at the key points in the color images are first detected and matched. Then, the common 3D points in two point clouds can be identified, based on the locations of the matched features in the color images plus their corresponding depth values. One example of these methods is shown in Figure 2b, where the 2D visual features are detected and matched with an oriented BRIEF (ORB) descriptor (Rublee et al. 2011) (Figure 2c). Based on the test results, it is found that the scene registration with the ORB descriptor is faster, but the registration with the FPFH descriptor is not affected by the lighting conditions in the environments.

(a) Scene registration using FPFH   (b) Matching in 2D

(c) Progressive scene registration using OBR
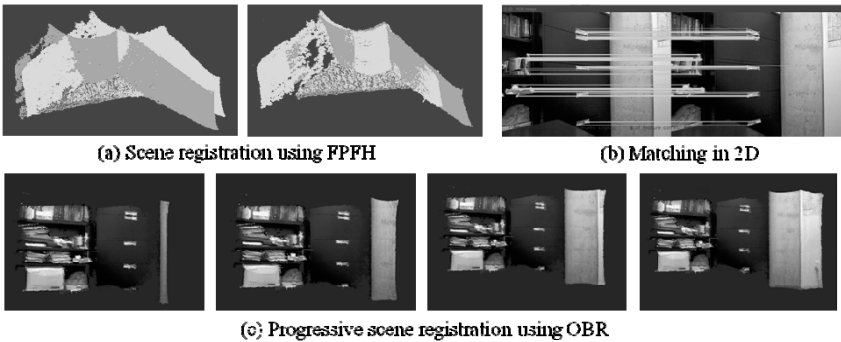
Figure 2. Scenes registrations with FPFH and OBR descriptors .

Study III: 3D Building Elements Recognition

The registration is to merge and align the point clouds from multiple scans into one single point cloud. The merged point cloud only includes the points with 3D coordinates. It does not explicitly contain any semantic or high-level geometric information, which is critical for as-built modeling. Therefore, the recognition is necessary to classify the points in the cloud, so that the points belonging to the same building elements in the scene can be grouped.

This study evaluates whether it is possible to exploit the visual features in color images to facilitate the recognition of building elements in 3D. Specifically, the surfaces of building elements in color images are first recognized based on their unique visual patterns. Then, the surface areas are mapped into their corresponding 3D points. This way, the points in the clouds are classified. Figure 3 shows the recognition of a concrete column surface in 3D with this idea.
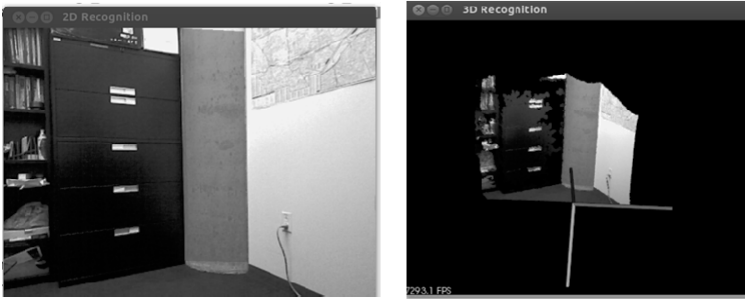
(a) Column surfaces recognition in 2D (b) Column surfaces recognition in 3D

Figure 3. 3D Concrete Column Surfaces Recognition.

CONCLUSIONS AND FUTURE WORK

3D as-built building models are useful, but the automatic generation of such models is challenging. This is especially true for modeling indoor environments, where the limited lighting conditions, large number of interior partitions, and prevalence of texture-poor building elements may fail the current automatic modeling process. The recent development in the RGB-D cameras provides a new way for as-built modeling. In order to understand the potentials of the RGB cameras, this paper evaluated the Microsoft® Kinect, one of common RGB cameras through three pilot studies. In the first study, the sensing data captured by the Kinect were compared with the data from a Total Station. It was found that the accuracy of the Kinect sensing data was at the centimeter level. The second study tested the registration of indoor scenarios by detecting and matching spatial and visual features separately, and the third study investigated the recognition of building elements through the fusion of spatial and visual data captured by the Kinect. Future work will focus on the development of an as-built modeling framework based on the findings in this paper.

ACKNOWLEDGEMENT

REFERENCES

Adan, A., Xiong, X., Akinci, B., and Huber, D. (2011). "Automatic creation of semantically rich 3D building models from laser scanner data," Proc. of ISARC 2011, Seoul, Korea
ClearEdge3D. (2012). "3D imaging services reduces MEP workflow on Pharma project by 60% using EdgeWise Plant 3.0." < http://www.clearedge3d.com/>

Durrant-Whyte, H. and Bailey, T. (2006). "Simultaneous localization and mapping," IEEE Robotics and Automation Magazine, 13(2): 99 – 110.

Fioraio N. and Konolige, K. (2011). "Realtime visual and point cloud slam.", In Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf. (RSS), 2011.

Foltz, B. (2000). "Application: 3D laser scanner provides benefits for PennDOT bridge and rockface surveys." Prof. Surv., 20(5): 22–28.

Furukawa, Y. and Ponce, J. (2010). "Accurate, dense, and robust multi-view stereopsis," IEEE Trans. on Pattern Analysis and Machine Intelligence, 32(8): 1362-1376.

Furukawa, Y., Curless, B., Seitz, S., and Szeliski, R., (2009). "Reconstructing building interiors from images," Proc. of ICCV 2009, Kyoto, Japan.

Geiger, A., Zeigler, J. and Stiller, C. (2011). "Stereoscan: Dense 3D reconstruction in real-time," Proc of IEEE Intelligent Vehicles Symposium, BadenBaden, Germany, 2011.

Haddad, N. and Ishakat, F. (2007). "3D Laser Scanner and Reflectorless Total Station: A Comparative Study of the Slots of El-Khazneh at Petra in Jordan, In: XXI International CIPA Symposium, Oct. 1-6, Athens, Greece.

Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D., (2010). "RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments", In: Proc. of International Symposium on Experimental Robotics, Delhi, India.

Izadi, S., et al. (2011). "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera.", ACM Symposium on User Interface Software & Technology.

Khoshelham K, and Elberink S. (2012). "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications." Sensors. 2012; 12(2):1437-1454

Okorn, B., Xiong, X., Akinci, B. and Huber, D. (2010). "Toward automated modeling of floor plans," Proc. of 3DPVT 2010, Paris, France.

Pollefeys, M., et al. (2008). "Detailed real-time urban 3D reconstruction from video". International Journal of Computer Vision, 78(2): 143 - 167.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G. (2011). " ORB: an efficient alternative to SIFT or SURF." In: Proc. of ICCV 2011: 2564-2571

Rusu, R. (2010). " Semantic 3D object maps for everyday manipulation in human living environments." Ph.D. thesis, Technische University.

Snavely, N., Seitz, S., and Szeliski, R. (2006). "Photo tourism: exploring image collections in 3D," Proc of SIGGRAPH 2006, Boston, MA.

Tang, P., Huber, D., Akinci, B. Lipman, R., and Lytle, A. (2010). "Automatic reconstruction of as-built building information models from laser-scanned point clouds: a review of related techniques," Auto. in Constr., 19(7): 829-843,

Whelan, T., Kaess, M., Fallon, M., Johannsson, H., Leonard, J., and McDonald, J. (2012). "Kintinuous: Spatially Extended KinectFusion"

Xiong, X. and Huber, D. (2010). "Using context to create semantic 3D models of indoor environments," Proc. of BMVC 2010, Aberystwyth, UK.

On Effective Text Classification for Supporting Job Hazard Analysis

N. W. Chi[1] , K. Y. Lin[2] , and S. H. Hsieh[3]

[1]Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan; PH +886-2-3366-3366#55202; email: naiwenchi@caece.net
[2]Department of Construction Management, University of Washington, Seattle, WA 98195-1610, USA; PH +1-206-616-1915; FAX +1-206-685-1976; email: kenyulin@uw.edu
[3]Department of Civil Engineering, National Taiwan University, Taipei 10617, Taiwan; PH +886-2-3366-4313; FAX +886-2-2363-1558; email: shhsieh@ntu.edu.tw

ABSTRACT

Construction industry is known to be dangerous and is featured by fatal occupational hazards. Job Hazard Analysis (JHA) is a common approach to mitigate and control these occupational hazards. It analyzes major tasks in a construction activity, identifies all the potential task related hazards, and suggests means to reduce or avoid each potential hazard.

Because every project is unique, an effective JHA must consider the conditions specific to the project instead of reusing previous JHAs directly. To this end, JHA requires the participation of experienced construction practitioners and becomes a time-consuming and brain-draining task. While expert involvement is necessary during JHA development, the authors would like to explore the possibility of leveraging existing construction safety resources in order to reduce the required human efforts.

This paper presents an approach based on text classification to support the automation of JHA. It uses the CPWR construction solutions database as an example to demonstrate how text classification can be applied to match the database documents with predefined safety violation scenarios and identify potentially useful safety approaches. This paper also discusses how different strategies were tested to optimize the effectiveness of the proposed text classification approach. The results indicate that although some of the classification strategies cannot make obvious progress on the effectiveness, the effectiveness of the text classification without optimization is good enough to support JHA.

INTRODUCTION

Construction industry is an industry that has high occupational hazards (Sorock, Smith et al. 1993). Comparing with other industries, there are more fatal hazards in construction, such as fall from heights, electrical shocks, and struck-by accidents. Therefore, safety management is always an important issue in construction. In the U.S., Job Hazard Analysis (JHA) is a popular solution for accident prevention. It identifies the potential hazards associated with the construction activities that are

taking place, and then recommends safety approaches to mitigate each of the potential hazards. Despite JHA's importance, it is not an easy task. Due to the wide spectrum of site conditions, job constrains, and work environments, construction projects are never the same, meaning that JHA must be recompiled instead of simply using existing JHA documents without making any changes.   The authors argue that developing a semi-automated approach for extracting useful information from existing construction safety text resources could provide a good initial reference for JHA.

In this research, text classification (TC) is applied to implement the semi-automated approach. TC is the task to assign documents into pre-defined categories and the proposed semi-automated approach is to "classify construction safety documents into existing unsafe scenarios" based on the activities and tasks involved. According to Wang and Boukamp (2011), the procedure of JHA is to find safety approaches under specific conditions that are featured by existing activities, tasks, and potential hazards. Therefore, construction safety documents which contain these important elements (i.e. activities, tasks, hazards) may have the potential to further infer safety approaches for JHA. In this paper, the authors will focus on the strategies that can yield better effectiveness on text classification over the construction safety documents. Specifically, the CPWR construction solution database (CPWR 2012) is used as an example to illustrate how optimization strategies could be obtained to improve the TC effectiveness.

METHODOLOGY: TEXT CLASSIFICATION ON CONSTRUCTION SAFETY DOCUMENTS

As mentioned in the introduction, this research performs TC on existing construction safety documents to find potential safety approaches for JHA. The procedure is shown in Figure 1. The documents are collected from two different resources. After that, they are combined and modified according to several different strategies that may achieve better TC effectiveness. Finally, the modified documents are classified and the TC effectiveness is calculated for evaluating the significance of the strategies.

THE PROCESS FOR CLASSIFICATION

The procedures illustrated in Figure 1 are used to guide the classification of construction safety documents. Because the procedures are designed to work with a wide range of document sources, some level of document preparation and enhancement is necessary in order to achieve better TC effectiveness. For this reason, the authors have used the construction safety standards from the U.S. Occupational Safety and Health Administration (OSHA) to enhance the CPWR solution database documents before classifying the documents. Properties of the two document collections are briefly discussed as that follows.
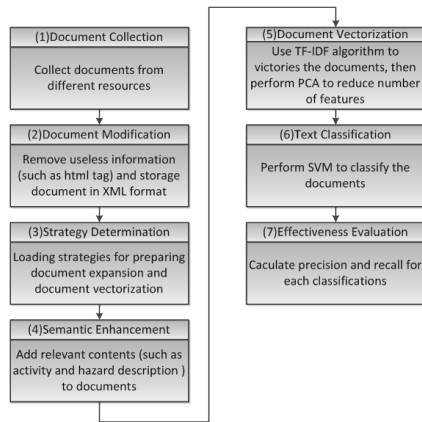
Figure 1 The process for construction safety document classification

THE CPWR CONSTRUCTION SOLUTION DATABASE AND THE OSHA STANDARDS

CPWR, The Center for Construction Research and Training, created the construction solution database which provides rich information on activities, tasks, and hazards that are involved in construction projects. These elements can describe the different work scenarios in construction. Each construction safety violation scenario in CPWR construction solution database can be treated as an independent document which contains the essential elements of activities, tasks, potential hazards, and safety approaches. Up to now, there are 15 types of construction activities (e.g. roofing) and 16 types of potential hazards (e.g. fall from height). However, not every type of activity or hazard is defined as a classifying category, because some of them do not contain sufficient instances to serve as training data. This research finally chooses 15 activities and 13 hazards to perform two classification scenarios. (i.e. the same documents will be classified twice according to two different point of views: activity and hazard)

The U.S. OSHA, mandates safety regulations for many industries, including construction. Its construction safety standards have twenty-eight subparts, for example "welding and cutting", "electrical", and "fall protection". Each subpart represents an important construction safety issue and there are many safety rules under each subpart. In this research, OSHA standards are applied to enrich the semantics of the CPWR solution database documents and to increase the training data size for better classification effectiveness.

EXPERIMENTS: TEXT CLASSIFICATION STRATEGIES

After determining the procedure, this research applies ten classification strategies to perform classification over the documents in the CPWR construction solution database as Table 1 shows. These strategies can be mainly divided into three groups: feature selection strategies, semantic expansion strategies, and co-reference expansion strategies. They try to add meaningful contexts and descriptions into the documents, or select important text feature to perform TC. Each strategy is controlled

by four different parameters (as the last four columns in Table 1). The details of the three groups of strategies will be further discussed in the next three sections. As discussed in the previous section, the documents in the CPWR construction solution database are to be classified based on the activities and hazards they involve. Therefore in the following discussion, the classification effectiveness (precision and recall) indicate how well the documents could be assigned to the type of activities (or hazards) they are related to.

Table 1 Classification Strategies

No.	strategy name	strategy description	activity description expansion	hazard description expansion	No. of standards expanded	No. of features for classification
1	baseLine	use original document without any modification	x	x	0	50
2	expandA	add activity description into each document	o	x	0	50
3	expandH	add hazard description into each document	x	o	0	50
4	expandAH	add both activity and hazard description into each document	o	o	0	50
5	pca25	classify document by the top 75 features	x	x	0	25
6	pca75	classify document by the top 75 features	x	x	0	75
7	pca100	classifying document by the top 100 features	x	x	0	100
8	osha5	add top 5 relevant osha standards to each document	x	x	5	50
9	osha10	add top 10 relevant osha standards to each document	x	x	10	50
10	osha15	add top 15 relevant osha standards to each document	x	x	15	50

FEATURE SELECTION STRATEGIES

Joachims (1998) indicates that TC has the characteristic of huge feature space and too many features may cause low effectiveness or even failure on automatic classification. As a result, this research adopts the Principal Component Analysis (PCA) (Abdi and Williams 2010)to perform feature selection. PCA not only can reduce the number of features but also rank their importance by their Eigen values. Therefore, if the new feature space is still too large to perform automatic classification, one can still only choose some of the "most important features" (i.e.

principal components) for the automatic classification. This research chooses 50 features as the baseline (baseLine), and compares it with 25 features (pca25), 75 features (pca75), and 100 features (pca100). Part of the classification results are illustrated in Table 2 (based on activities) and Table 3 (based on hazards).

Table 2 Activity-based classification results for various feature selection strategies

scenario information		baseLine		pca25		pca75		pca100	
id	activity name	precision	recall	precision	recall	precision	recall	precision	recall
3	Electrical	0.425	0.607	0.792	0.679	0.618	0.750	0.952	0.714
4	Excavation and Demolition	0.786	0.458	0.727	0.333	0.947	0.750	0.739	0.708
6	Heavy Equipment	0.286	0.429	0.167	0.143	1.000	0.214	0.813	0.929
11	Reinforced Concrete	0.571	0.625	0.405	0.531	0.564	0.688	0.786	0.688
13	Roofing	1.000	0.625	1.000	0.438	0.789	0.469	0.690	0.625
15	Structural Steel	1.000	0.885	0.889	0.615	0.828	0.923	0.828	0.923

Table 3 Hazard-based classification results for various feature selection strategies

scenario ID		baseLine		pca25		pca75		pca100	
id	hazard name	precision	recall	precision	recall	precision	recall	precision	recall
16	Cuts and punctures from tools or materials	1.000	1.000	1.000	0.786	1.000	0.786	1.000	0.786
17	Fall from heights	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
19	Manual materials handling	0.571	0.462	0.852	0.885	0.923	0.923	0.925	0.942
24	Stressful hand and wrist activity	1.000	0.982	1.000	0.964	0.982	0.964	0.982	0.964
26	Kneeling and Squatting	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
27	Construction dust	0.565	1.000	0.545	0.923	0.647	0.846	0.647	0.846

SEMANTIC EXPANSION STRATEGIES

In information retrieval (IR) field, query expansion, meaning appending relevant terms such as synonyms to their query terms, is a general approach to enhance the search effectiveness (Buckley, Salton et al. 1994; Huang, Chien et al. 2003). When it comes to TC, appending more semantic information to training data may also achieve better effectiveness. In the CPWR construction solution database, there are not only safe solutions for hazardous scenarios but also rich information and descriptions about the activities and hazards. Since the TC will be performed base on the activity and hazard concepts, this research tries to append "only the activity

descriptions (expandA)", "only the hazard descriptions (expandH)", and "both the activity and hazard descriptions (expandAH)" to the training data. Due to the manuscript length limitation, the classification result tables are not included here.

CO-REFERENCE EXPANSION STRATEGIES

Similar to the discussion in Section 3.2, appending relevant semantic text contents into a document may achieve better classification effectiveness. Besides expanding activity and hazard descriptions, one additional semantic expansion strategy is to append the contents of "references" to documents. If two documents refer to the same reference, they are more likely to be related to each other. Based on this principal, this research applies information retrieval techniques to identify the OSHA standards (as the document collection) that are related to the different activity and hazard combination scenarios (as the query terms) in the CPWR solution database. After every CPWR scenario is appended using relevant OSHA standards, the research then evaluates how the co-reference effect (i.e. two CPWR documents referring to the same OSHA standards) enhances the effectiveness of TC. Three different expansion strategies are generated as a result, including appending the CPWR documents with the top 5, 10, and 15 relevant OSHA standards. Again, due to the manuscript length limitation, the classification result tables are not included here.

RESULTS AND DISCUSSIONS

Wilcoxon Signed Rank Test (Wilcoxon 1945) provides a promising way to evaluate the TC performance for each strategy discussed in Section 3. When two strategies are compared with each other, there are four points of views: precision under activity classification, recall under activity classification, precision under hazard classification, and recall under hazard classification. Table 4 shows the results of the Wilcoxon Signed Rank Test on the activity classification. As for the hazard classification, since all the null hypothesis is rejected, the results are not shown in table due to the length limitation. From Table 4, the authors conclude that:

1. No one strategy is significantly better than another strategy when classifying the CPWR documents into different hazard groups. However, as exemplified by Table 4, all the classification strategies including the baseline strategy achieve a fair level of precision and recall. It means that we may not need to apply any modification to the text.
2. The osha15 strategy is the best strategy among all the co-reference expansion strategies. It is possible that osha5 and osha10 are not significantly better than the baseLine strategy because the appended information is still insufficient to make a difference during TC.
3. The expandA (expanding only the description of activity) is the best semantic expansion strategy when classifying CPWR documents into different activity groups. The strategy also seems to contribute more on recall than precision.
4. The pca100 strategy (choosing the top 100 features among the principal components) is the best feature selection strategy. Although too many features may disturb the effectiveness of TC, the benefit of selecting sufficient features to

represent the documents seems to out weight the potential disturbance.

Table 4 Wilcoxon Signed Rank Test results over activity classification

Hypothesis		Precision			Recall		
H0	H1	W	critical value	reject H0	W	critical value	reject H0
baseLine>= osha5	baseLine< osha5	5	26	FALSE	0	26	FALSE
baseLine>= osha10	baseLine< osha10	17	26	FALSE	13	21	FALSE
baseLine>= osha15	baseLine< osha15	18	17	TRUE	16	17	FALSE
osha5>= osha10	osha5< osha10	81	21	TRUE	52	11	TRUE
osha5>= osha15	osha5< osha15	85	21	TRUE	81	26	TRUE
osha10>= osha15	osha10< osha15	47	26	TRUE	42	21	TRUE
baseLine>= expandA	baseLine< expandA	51	26	TRUE	46	30	TRUE
baseLine>= expandH	baseLine< expandH	31	26	TRUE	14	21	FALSE
baseLine>= expandAH	baseLine< expandAH	0	30	FALSE	0	30	FALSE
expandH>= expandA	expandH< expandA	43	26	TRUE	62	26	TRUE
expandA>= expandAH	expandA< expandAH	15	26	FALSE	32	30	TRUE
expandH>= expandAH	expandH< expandAH	35	26	TRUE	64	21	TRUE
pca25>= pca75	pca25< pca75	79	21	TRUE	78	17	TRUE
pca25>= pca100	pca25< pca100	72	21	TRUE	120	30	TRUE
baseLine>= pca25	baseLine< pca25	15	30	FALSE	3	26	FALSE
baseLine>= pca75	baseLine< pca75	46	17	TRUE	63	26	TRUE
baseLine>= pca100	baseLine< pca100	40	21	TRUE	54	26	TRUE
pca75>= pca100	pca75< pca100	33	17	TRUE	32	26	TRUE

CONCLUSIONS

In this research, the authors simulated the logic of JHA that "using the safety violation scenario to find appropriate safety approaches", transformed the logic into a text classification problem, demonstrated the semi-automated procedure for matching the documents and the safety violation scenarios, and finally tried to optimize the TC effectiveness by applying different TC strategies. In practical use, since we already have sufficient training data in the CPWR construction solution database, if there are other construction safety documents (such as JHA documents, safety standards, and so on.), we can apply similar procedure to find potential safety approaches for JHA. This research also illustrates how to find better text classification strategy for specific document collection. Although these parameters may change with different types of document set, they can be still re-performed under different conditions. In this research, both the training data and the test data comes from the same source, so the good effectiveness on text classification may be caused by their similarities, such as

the overlapping knowledge scope, the same terminologies, and so on. In the future, the authors will try to extent the approach over different types of construction safety documents.

REFERENCE

Abdi, H. and Williams, L. J. (2010). "Principal component analysis." Wiley Interdisciplinary Reviews: Computational Statistics 2(4): 433-459.

Buckley, C., Salton, G., Allan, J., and Singhal, Amit (1994). "Automatic Query Expansion Using SMART: TREC 3". TREC.

CPWR (2012). "Construction Solutions." Retrieved Dec.24, 2012, from http://www.cpwrconstructionsolutions.org.

Huang, C. K., Chien, L. F., and Oyang, Y. J. (2003). "Relevant term suggestion in interactive web search based on contextual information in query session logs." Journal of the American Society for Information Science and Technology 54(7): 638-649.

Joachims, T. (1998). "Text categorization with Support Vector Machines: Learning with many relevant features." Machine Learning: ECML-98. Springer Berlin / Heidelberg. 1398: 137-142.

OSHA (2012). "Occupational Safety and Health Administration." Retrieved Dec.24, 2012, from http://www.osha.gov/.

Sorock, G. S., Smith, E. O., and Goldoft, M.(1993). "Fatal occupational injuries in the New Jersey construction industry, 1983 to 1989." Journal of Occupational Medicine 35(9): 916-921.

Wang, H. H. and Boukamp, F. (2011). "Ontology-Based Representation and Reasoning Framework for Supporting Job Hazard Analysis." Journal of Computing in Civil Engineering 25(6): 442-456.

Wilcoxon, F. (1945). "Individual comparisons by ranking methods." Biometrics 1: 80-83.

Active Dimensional Quality Assessment of Precast Concrete using 3D Laser Scanning

M. K. Kim[1,2] , H. Sohn[1] and C. C. Chang[2]

[1]Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 291 Daehak-Ro; PH (82) 42-350-3665; FAX (82) 42-350-3610; email: {joekim, hoonsohn}@kaist.ac.kr
[2] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, Clear Water Bay, Kowloon; PH (852) 2358-7201; FAX (852) 2358-1534; email: cechang@ust.hk

ABSTRACT

Quality assessment of precast concrete panels is an important factor that affects overall quality of construction. As prefabrication becomes popular at construction sites, demands for automatic and accurate inspections of the dimensions of the precast concrete panels have increased. Current techniques for measuring the dimensions of precast concrete panels, however, heavily rely on qualified inspectors, and are time and labor demanding. To overcome these limitations, a dimensional measurement technique for precast concrete panels is proposed using a 3D laser scanner. For autonomous implementation of the dimensional measurement, a new feature extraction algorithm of extracting only boundary points of the precast concrete panels is developed. To increase the measurement accuracy, a compensation model is employed to account for the dimension losses caused by an intrinsic limitation of laser scanners. This study focuses on measurements of length, width and squareness of precast concrete panels in a non-contact and speedy manner. Experimental validations on actual precast slabs are conducted to verify the applicability of the proposed dimensional measurement technique for precast concrete panels.

INTRODUCTION

Precast concrete panels are a popular type of construction components nowadays. As compared to conventional in-situ-cast concrete components, precast concrete panels have more consistent properties and can reduce the construction time and cost (Sacks et al. 2004). Prefabrication also contributes to a tidier and safer working environment on-site when compared to conventional construction. Although with these advantages of precast concrete construction, the performance of the finished product is heavily dependent on the quality of the product itself so that quality inspection on precast concrete components is an essential task for success in precast constructions.

Currently, the quality inspection of precast products is evaluated manually by certified inspectors. The inspectors commonly follow the guidelines provided by International Organization for Standardization (ISO 9001, 2008) and Precast/Prestressed Concrete Institute (PCI, 2000). For the visual inspection on precast concrete, however, inspectors heavily rely on contact-type measurement device such as straightedge and profilometer (Latimer et al. 2002), which is subjective and time consuming. In order to overcome these limitations, several researches have been investigated on quality inspection of concrete structures by using smart sensors (Hutchinson et al. 2006, Zhu and Brilakis 2010, Tang et al. 2011, Akinci et al. 2006). However, there has been not investigated on dimensional quality assessment for precast concrete panels in a speedy and automatic manner.

In this paper, we present a new dimensional quality assessment technique for precast concrete panels by using a 3D laser scanner. A novel feature extraction algorithm called 'Vector-Sum algorithm' is developed based on the unique characteristic of the laser scanner. To increase the measurement accuracy, a compensation model is employed to account for the dimension losses caused by an intrinsic limitation of laser scanners. To the authors' best knowledge, this is the first study that has proposed an autonomous dimensional measurement technique for precast concrete panels. This paper is organized as follows. First, methodology of the proposed precast concrete dimensional measurement technique is formulated. Then, experimental validations on real precast slabs manufactured for this study are followed to verify the applicability of the proposed precast concrete dimensional measurement technique. Finally, this paper concludes with a brief summary and discussions for future work.

METHODOLOTY

(a) Overall scheme   (b) Laboratory test configuration
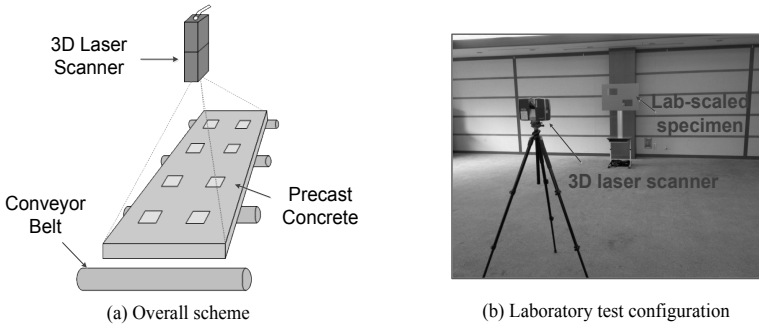
Figure 1. The proposed precast concrete dimensional quality assessment system

Figure 1 shows the overview of the proposed dimensional quality assessment system for precast concrete panels. The system is intended to provide semi-autonomous procedure, starting from data acquisition provided by the laser scanner to classification of the quality of precast concrete products. It is assumed that the manufactured precast concrete sample is delivered via the conveyor belt and positioned at a designated location. Then, the laser

scanner located above the center of the precast concrete scans the whole area of the precast concrete in a single scan. With a single scan, rapid and automated dimensional inspection of the precast concrete panels is possible. After the inspection is completed, decision on the sample whether it is acceptable or not is made based on the tolerance of the quality assessment criteria specified in PCI tolerance manual (PCI, 2000). Figure 1(b) shows the laboratory test configuration designed for validation of the proposed dimensional quality assessment technique and for better understanding of the following section.

(1) Acquisition of 3D laser scan data

(2) Selection of corner points on range image

(3) Coordinate transformation from scanner to object

(4) Filtering of background data

(5) Surface fitting & projection into the fitted plane

(6) Edge and corner extraction
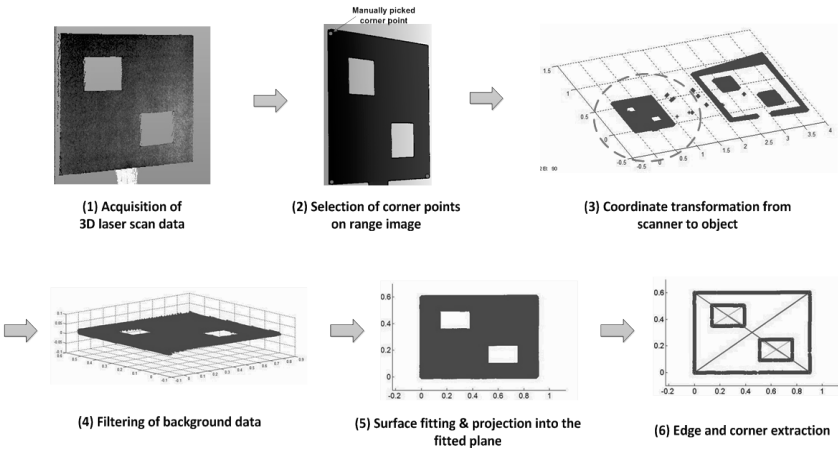
Figure 2. The procedure of the proposed semi-autonomous dimensional assessment technique

The overall procedure of the proposed semi-autonomous dimensional quality assessment technique for precast concrete panels is illustrated in Figure 2. (1) It starts with acquisition of the geometric information of a precast slab through a 3D laser scanner (2) Then, selection of three points near each corner of the target is manually performed from the range image. Here, the range image is a 2D image that each pixel of the image presents the distance value between the scan point and the origin of the laser scanner. (3) After that, coordinate transformation is conducted from scanner coordinate system to object coordinate system for better recognition of the geometric information and rapid analysis for post processes. (4) Sequentially, elimination of unwanted background scan points is performed by setting a boundary region of the transformed scan data. (5) Then, surface fitting using the filtered scan points is conducted and projection of the points into the fitted surface is implemented. With this projection process, dimension reduction of the data is made from three-dimensional space to two-dimensional space, resulting in significant reduction of computation time for post data analysis. (6) As the last step, edge and corner extractions are followed. A novel algorithm, called 'Vector-Sum', is developed to extract only edge points along horizontal and vertical edge lines of the precast slab. More details of the Vector-Sum algorithm are presented in the following paragraph. Next, the dimensions of the precast slab are calculated by employing a dimensional compensation model which compensates

for dimension loss caused by an intrinsic limitation of the laser scanner. Also, details of the dimensional compensation model are also described later in this paper. Finally, by comparing the estimated dimensions with those of design model, classification on whether the monitored precast slab is accepted or not is performed.
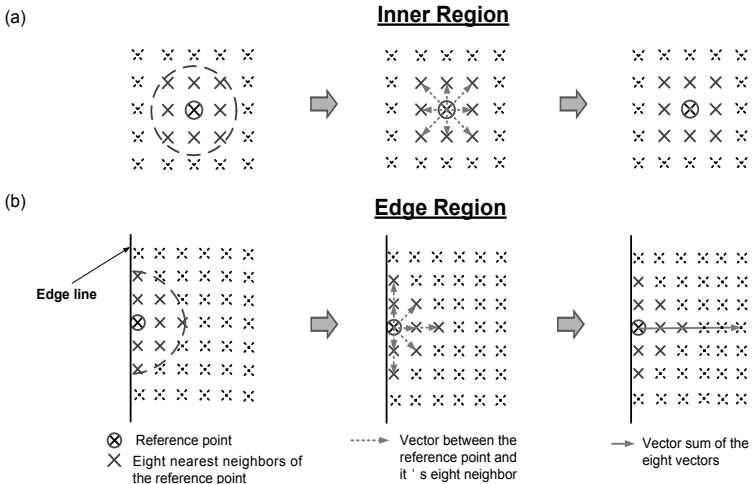
Figure 3. Vector-Sum algorithm for edge point detection: (a) Implementation in an inner region of precast concrete; (b) Implementation in an edge region

Figure 3 describes the working principle of the Vector-Sum algorithm developed by the authors. The aim of the Vector Sum algorithm is to extract pure edge points located along the edge lines of the precast concrete for accurate dimension inspection. This algorithm is basically assumed that the scanned data of the precast concrete is neatly aligned along both horizontal and vertical directions. For this algorithm, k-nearest neighbors for a point is employed for edge point extraction. If a scan point is selected as reference point, the k-nearest neighbor points of the reference point can be automatically extracted based on Euclidean distance. Once the k-nearest neighbors of the reference scan point are selected, k number of vectors is created by connecting the reference point and each neighbor point of the reference point. Then, summation of the formulated k vectors is conducted. In this study, the number of k is determined as eight because the number of eight maximizes the performance of extracting edge points with minimum computation cost. From the implementation of the Vector-Sum algorithm, distinction between the edge points and the non-edge points is clearly performed. It means that, in the case of the inner region of the precast slab as shown in Figure 3(a), the magnitude of the vector summation is theoretically zero, whereas the magnitude of the vector summation is five times the spacing interval between two adjacent scan points in the edge region of the precast slab as shown in Figure 3(b). This indicates that extraction of only edge points of precast concrete can be realized and it contributes to accurate dimension estimation of the precast slab.

(a)   (b)

Mixed pixel phenomenom
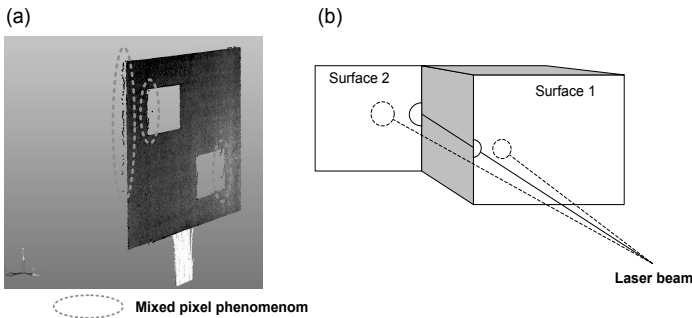
Figure 4. Mixed-pixel phenomenon: (a) An example of mixed-pixels from the obtained scan data; (b) Cause of mixed-pixels

Although laser scanners generally are regarded as accurate and reliable means of acquiring spatial information about real objects, certain environmental and sensing conditions of laser scanning can lower the accuracy. One such condition is known as mixed-pixel effect (Hebert et al. 1991, Mills et al. 2004). Figure 4 shows the mixed-pixel phenomenon. In Figure 4(a), it can be seen that some points in the edge region of the object deviate from the object surface. The mixed pixel phenomenon is occurred when the laser beam lies partially on two surfaces having difference distances from the laser scanner as shown in Figure 4(b). In this situation, the laser beam is reflected by the two surfaces, and the encorder of the laser scanner receives a mixture of the two signals, resulting in inaccurate range measurements of the laser scanner. To prevent possible difficlties in post data analysis, these deviated points are commonly removed from the scan data by setting bounds in a region of interest. However, the removal of the mixed pixels at the edge region prevents precise measurement of the object's dimensions such as length and width. To avoid this problem, in this study, an edge loss model is employed to compensate for dimension loss of the precast concrete (Tang et al. 2010).

VALIDATION

To examine the effectiveness of the proposed dimensional inspection technique, experimental tests were conducted on actual precast slabs. The overall test configuration and the investigated precast slabs are shown in Figure 5. The bottom surface of the precast slab was fixed on a concrete structural wall and was scanned by the scanner. The position of the laser scanner was determined and fixed at 2.5 m height from the ground and 10 m distance from the precast slab. After the determination of the scan position, scans on the two modules were conducted with three angular resolution cases of 0.009°, 0.018° and

0.036°. The scanning times for each angular resolution case were measured approximately at 5, 3 and 1 minutes, respectively. For this experiment, two different actual
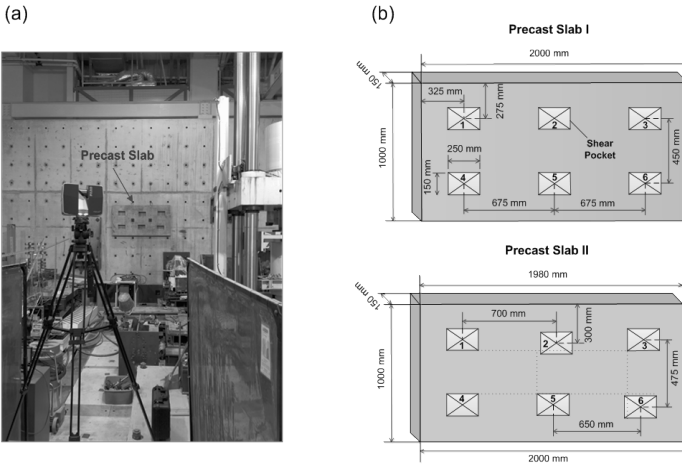
Figure 5. Experimental configuration on actual precast slab: (a) Test configuration; (b) Precast slab specification

precast slabs were used, one is assumed to be a normal condition precast slab (Precast slab I) and the other one is intended for an abnormal condition precast slab (Precast slab II) with dimensional errors over the tolerance as shown in Figure 5 (b). In each precast slab module, there are six identical rectangular shear pockets with dimensions of 250 mm x 150 mm. For the abnormal precast slab, three abnormalities are intentionally made, which are the 20 mm short upper horizontal dimension (1980 mm) of the precast slab and the position shifts of the shear pockets numbered as 2 and 6 with 250 mm x 250 mm in both horizontal and vertical directions. The reduction of the upper horizontal dimension of the precast slab was intended to recognize not only the dimension error of the precast slab but also squareness error of the precast slab. The squareness error is defined as the difference between the biggest and shortest dimensions of the longer sides of a precast component [PCI 2000]. As applied to this study, the longer side of the precast slab is the horizontal dimension such that squareness error of the precast slab is the dimension gap between the upper and lower horizontal lengths. Figure 6 shows the edge and corner extraction results for two actual precast concrete slabs. The result in Figure 6 was obtained from the case of angular resolution of 0.009°. Note that similar results were obtained for the cases of angular resolution of 0.018° and 0.036°, but they are not reported here due to the limited space. As expected, the implementation of Vector-Sum process provides a great performance in extraction of only the edge points in both the two precast slabs. The corners of the precast slabs and six shear pockets for each slab are extracted by conducting line fittings on the edge points. Then, using the dimensional compensation model, errors of dimension, position and squareness for each precast slab are estimated.
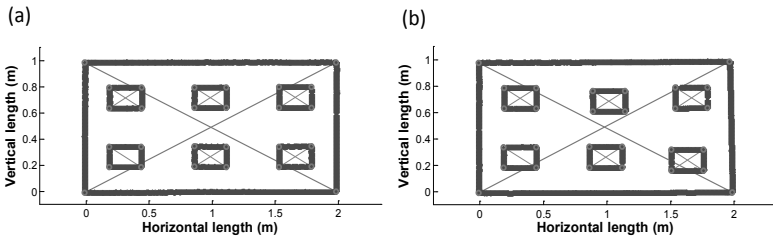
Figure 6. Edge and corner extraction results: (a) Precast slab I; (b) Precast slab II

Table 1. Dimensional errors for the actual precast slabs

Slab	Angular resolution (°)	Dimensional Error (mm)		
		Dimension	Position	Squareness
Precast slab I	0.009	1.3	1.7	0.0
	0.018	1.2	1.7	0.0
	0.036	2.0	2.4	0.0
Precast slab II	0.009	1.3	2.0	3.5
	0.018	1.7	3.4	1.2
	0.036	3.9	1.7	5.5

Table 1 summarize the dimensional estimation results for the two actual precast slabs. For the Table 1, the dimensional errors calculated in both horizontal and vertical directions are averaged such that one dimensional error is presented for each angular resolution. For the dimension (length and width) error results, an accuracy of 97.6 %, having 41 cases out of 42 cases, were within the tolerance (6 mm) for precast slab I and 92.8 % for precast slab II. For the position errors of six shear pockets for precast slab I and II, similar results were obtained compared to the dimension error results, having accuracy of 100 % and 91.6 % for precast slab I and II, respectively. For the precast slab II, a large position error having average of 7.3 mm were obtained in shear pocket 6. It turned out through a visual inspection that there is a 5 mm horizontal position shift of shear pocket 6 compared to the initial design. The authors speculate that this error comes from a human error during the design stage for precast slab II. For the squareness error of the precast II, the intended squareness errors were successfully detected with level of within the tolerance. From these results, it can be concluded that the proposed dimensional measurement technique is robust and reliable for dimensional quality assessment of precast concrete panels.

CONCLUSION AND FUTURE WORK

This paper describes a new dimensional measurement technique for automated quality assessment of precast concrete using a 3D laser scanner. The current manual inspection system on precast concrete needs a new paradigm where accurate quality control can be instantaneously performed without human efforts. To meet this demand, a novel edge point

extraction algorithm called Vector-Sum algorithm was developed based on the unique characteristics of the scan data. A dimensional compensation model was also employed to enhance the dimension measurement accuracy for precast concrete panels. Experimental tests on actual precast slabs were conducted to demonstrate the applicability of the proposed technique for precast concrete quality inspection. For the actual precast slab experiments, 95.5 % cases of the dimension and position measurements lie within the tolerance range. The results indicate that the proposed technique is able to accurately measure dimensions of precast slabs to levels within the predetermined tolerance. However, the applicability of the proposed dimensional inspection technique for precast concrete is limited to flat-shaped precast concrete components. Further investigation is underway to extend the proposed technique to more complex shaped precast concrete components.

REFERENCES

Sacks, R., Estman, C. M., and Lee, G. (2004). "Process model perspectives on management and engineering procedures in the precast/prestressed concrete industry." *J. Constr. Eng. Manage.*, 130(2), 206-215.

International Organization for Standardization (ISO). (2008). *ISO 9001: Quality Systems*, Geneva.

Precast/Prestressed Concrete Institute (PCI). (2000). *Tolerance Manual for Precast and Prestressed Concrete Construction*. Retrieved from http://www.quality-concrete.net/library/precast/precast_tolerance.pdf.

Latimer, D., Gujar, S., Garrett, J., Akinci, B., Thayer, S., and Paredis, C. (2002). "Running surface assessment technology review." *ICES Research Project Rep.*, Carnergie Mellon Univ., Pittsburgh.

Hutchinson, T. C., Chen, Z. (2006). "Improved image analysis for evaluating concrete damage." *J. Comp. Civil Eng.*, 20(3), 210-216.

Zhu, Z. and Brilakis I. (2010). "Machine vision-based concrete surface quality assessment." *J. Constr. Eng. Manage.*, 136(2), 210-218.

Akinci, B., Boukamp, F., Gordon, C., Huber, D., Lyons, C., and Park, K. (2006). "A formalism for utilization of sensor systems and integrated project models for active construction quality control." *Autom. Constr.*, 15(2), 124-138.

Heber. M. and Krotkv, E. (1991), "3-D measurement from imaging laser ladars: how good are they?.", IEEE/RSJ *International Workshop on Intelligent Robots and Systems*, IROS'91, Osaka, Japan, 1, 359-364.

Mills, J. and Barber, D. (2004), "Geomatics techniques for structural surverying.", *J. Survey. Eng.*, 130(2), 56-64.

Tang, P., Akinci, B. and Huber, D. (2009), "Quantification of edge loos of laser scanned data at spatial discontinuities.", *Autom. Constr.*, 18(8), 1070-1083.

Tang, P., Huber, D., and Akinci, B. (2011), "Characterization of Laser Scanners and Algorithms for Detecting Flatness Defects on Concrete Surfaces." *J. Comp. Civil Eng.*, 25(1), 31-42.

Automating the Task-level Construction Activity Analysis through Fusion of Real Time Location Sensors and Worker's Thoracic Posture Data

T. Cheng[1], J. Teizer[2,*], G.C. Migliaccio[3], and U. Gatti[2]

[1] Ph.D. Candidate, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N.W., Atlanta, GA, 30332-0355, E-Mail: tcheng9@gatech.edu
[2*] Ph.D., Associate Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N.W., Atlanta, GA, 30332-0355, E-mail: teizer@gatech.edu (corresponding author)
[3] Ph.D., Assistant Professor, Department of Construction Management, College of Built Environments, University of Washington, 120D Architecture Hall, Box 351610 Seattle, WA 98195-1610, E-Mail: gianciro@uw.edu
[4] Ph.D., Research Associate, Department of Construction Management, College of Built Environments, University of Washington, 120D Architecture Hall, Box 351610 Seattle, WA 98195-1610. E-mail: ucg@u.washington.edu

ABSTRACT

Knowledge of workforce productivity and activity is crucial for determining whether a construction project can be accomplished on time and within budget. As a result, significant work has been done on improving and assessing productivity and activity at task, project, and industry levels. Task level productivity and activity analysis are used extensively within the construction industry for various purposes, including cost estimating, claim evaluation, and day-to-day project management. Nevertheless, assessment of task level productivity and activity analyses are mostly performed through visual observations and after-the-fact even though studies have been performed to automatically translate the construction operations data into productivity information and to provide spatial information of construction resources for specific construction operations. This paper presents an original approach to automatically assess construction labor activity. Using data fusion of spatiotemporal and workers' thoracic posture data, the authors have developed a novel framework for identifying and understanding the worker's activity type over time automatically. This information is used to perform automatic work sampling that is expected to facilitate real-time productivity assessment.

INTRODUCTION

Several studies reported, productivity in the construction industry has been declining over the past decades (Arditi and Krishna 2000). These analyses, however, are based on assembled measures from multiple governmental agencies (e.g., Census

Value of Construction Put in Place) and do not regard the broader concerns regarding the accuracy of such productivity measures. Until today, the aggregated productivity performance is not measured for the most part due to the lack of suitable and sustainable approaches to accurately and automatically monitor the actual activity and work output (Goodrum and Haas 2002).

Workforce productivity is a major aspect in determining if a construction project can be accomplished on time and within budget, an effective and timely approach to productivity management is crucial to the success of construction projects and construction companies. An extensive literature on construction productivity has confirmed the importance of these concepts to the success of construction projects (Gouett et al. 2011).

The productivity of this industry can be assessed at three levels: task, project, and industry level. Task level focuses on single construction activities, such as structural steel erection or concrete placing. Task level productivity is used extensively within the construction industry. Different construction tasks are combined at the project level. The BLS, however, currently does not maintain a productivity index for any sector of the U.S. construction industry. Current practice strongly relies on historical production rates. Changes in workforce composition are expected to produce uncertainties in historically-based production estimates. Hence, there is a need for data collection and processing approaches that would produce real-time automated productivity assessment.

This paper presents an original approach to automatically assess labor productivity. Using data fusion of spatio-temporal and workers' thoracic posture data, the authors have developed a framework for identifying and understanding the worker's activity type over time. This information is utilized to perform automatic work sampling that is expected to facilitate real-time productivity assessment.

PRODUCTIVITY ASSESSMENT, SENSING, DATA FUSION TECHNIQUES

Several productivity assessment methods have been created and adopted within the construction industry. The work sampling technique has been widely used to understand the characteristics of a work process in industrial settings. It is implemented as an indirect method to measure activity level and productivity. Another example is activity analysis, which is the evolution of the practice of work sampling. An activity level is defined as the percentage of time that craft workers spend on a particular activity (CII 2012). It is represented by the direct work time rate. In fact, measuring work rate is not the same as measuring productivity. Compared to the traditional work sampling technique, activity analysis includes significantly more detailed observations (Gouett et al. 2011).

Even though several existing methods can generate useful information to improve construction activities, many of them present severe limitations, including being manually intensive, involving human judgment, and being ineffective in providing timely and accurate control data. Thus, it is reasonable to assume that automated productivity assessment methods can be beneficial for construction.

With the development of new information and sensing technologies, it is possible to provide a steady and reliable data stream of construction process. Vision, radio frequency, and other remote sensing technologies promise effortless data

recording. However, their limitation is still in real-time data processing and information feedback (Abeid and Arditi 2002; Teizer and Vela 2009; Shahandashti et al. 2011; Cheng et al. 2012). Ultra Wideband (UWB) technology, as a infrastructure sensor-intensive active Radio Frequency Identification (RFID) technology, allows recording location data of multiple resources (worker, equipment, material, etc.) in real-time while the tracking error in a harsh construction environment is less than one half of a meter (Cheng et al. 2011).

STUDY OBJECTIVES, SCOPE, AND METHODOLOGY

Data fusion applications span a very wide domain (Shahandashti et al. 2011). By integrating data from real-time location sensors (RTLS) and thoracic accelerometers this study attempts to continuously assess task activities of construction worker(s). The goal of this research is to automate the process of activity analysis by fusing information on body posture and positioning factors of repeated manual material handling activities in construction environments. The first objective is to automatically identify and characterize the various site geometries related to different activities including work zone, material zone, and rest zone. The second objective is to automatically measure the direct work time rate by computing the time lapse of both productive and non-productive activities including wrench, material, traveling, and rest times.
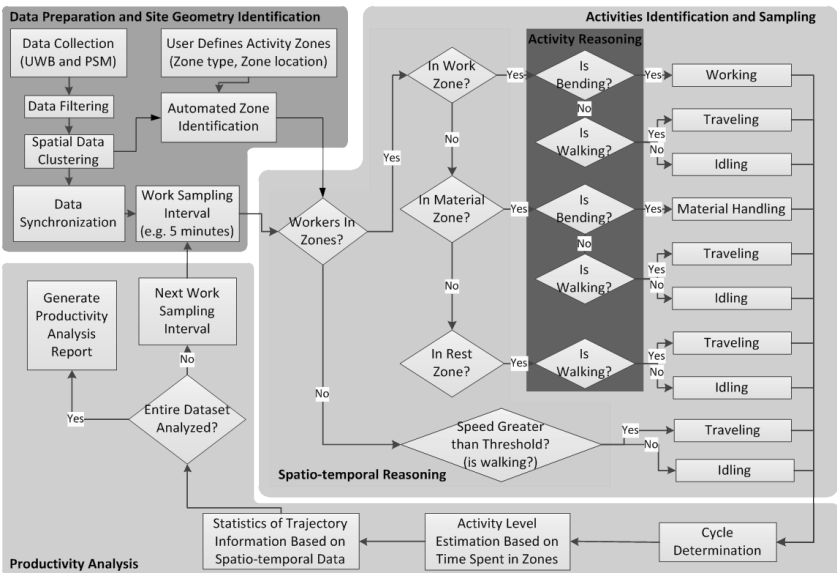
Figure 1: Automated activity analysis and productivity measurement by reasoning the workers' spatiotemporal information and posture status.

All tests were performed in a controlled study environment. Working activities occurred indoors and on the same elevation level. This study focuses only on the labor productivity measurement by integrating and analyzing the location data and thoracic posture information of the workers. The methodology (shown in Figure 1) consists of data preparation and site geometry identification, activity identification and sampling, and productivity analysis. The protocol of activity analysis is: (1) *Work*: the participant is performing the assigned construction task within a work zone (e.g., assembling floor tiles); (2) *Material*: the participant is handling material within a material zone (e.g., picking and placing); (3) *Travel*: the participant is moving between material, and rest zones; (4) *Idle*: any activity that is not work, material, or travel (e.g., staying inside a zone without performing any work).

DATA COLLECTION TEST BED & FUSION APPROACH FOR ANALYSIS

Data collection and filtering: A UWB localization system is utilized to monitor the real-time spatial and temporal information of the participants. UWB tags are placed on the helmets of the participants for location tracking purpose, as well as at static positions to identify special site geometry, including material, work, and rest zones. A Physiological Status Monitoring (PSM) system autonomously and remotely monitors the posture of the participants. A wearable 3-axial thoracic accelerometer generates the participant's posture measurement (bending angle) in Vector Magnitude Unit (VMU). Both data streams carry noises. The data synchronization technique and error analysis have been explained in the authors' previous work (Cheng et al. 2012).

Data fusion: Since PSM and UWB monitor the work activities on two different aspects and independent timelines, the attributes of the location tracking and posture information have to be fused. Fusing requires the data to be synchronized using probabilistic inference. A fuzzy representation is implemented to define the results of spatiotemporal reasoning and activity status reasoning. The spatio-temporal status is therefore described as "inside" or "outside of a zone", and the activity status is represented as "bending" and "walking". The likelihood function using Bayesian approach is computed at a specific reasoning status A_i at a given synchronization function, where S_{new}^k is new data from sensor k, S_{old}^k is old data from sensor k, and $P(A_{i|} \mid S_{old}^1 S_{old}^1)$ is the prior estimation in the previous data synchronization model.

$$\ell(A_i \mid S_{new}^1 S_{new}^2) = \frac{\prod_{k=1,2} P(A_i \mid S_{new}^k)\, P(A_i \mid S_{old}^1 S_{old}^2)}{\prod_{k=1,2} P(A_i \mid S_{old}^k)}$$

Activity sampling using spatiotemporal reasoning and activity status reasoning: The fused data is utilized to assess the work activities based on pre-defined data query rules. In order to achieve an accurate work activity assessment, it is crucial to define a set of proper activity categories, which must suit the need and the objective of the study and the feature of the work tasks that are being monitored. In this paper, the activities are sampled into four work categories: direct, material

handling, travel, and idle. Then, a two-step reasoning mechanism is applied: spatio-temporal reasoning and activity reasoning.

The fused data stream is firstly queried on the spatial and temporal aspect. The geometrical relationship between the participant's trajectories and the updated zone definition is checked. Three zone types are assigned: work zone, material zone, and rest zone. Further data is reasoned by the participant's thoracic posture data. Staying in a specific zone will not be identified as a corresponding activity unless a motion change of the participant's thoracic posture status is observed. The fused data is then classified by activity status: working, traveling, material handling, and idling.

Identification of direct work activity requires understanding of the behaviors of the participant. A criterion is set: the participant has to be present in the work zone and have a high posture angle. Activities in the work zone with up-right posture will be considered as either traveling or idling according to the participant's moving speed. Similarly, the material handling activities are identified through the posture status and movement of the participant inside and/or outside the specific zones.

Productivity analysis: As the activity type has been identified, the work cycle information (activity duration and sequencing) is based on start and end time of the activity. The activity level, which is described by the rate of direct work time versus total time (following equation), is automatically computed:

$$Estimated\ Direct\ Work\ Time\ Rate = \frac{Time\ of\ Identified\ Direct\ Work}{Total\ Time}$$

$$\geq Actual\ Direct\ Work\ Time\ Rate = \frac{Time\ of\ Direct\ Work}{Total\ Time}$$

EXPERIMENT AND RESULTS

The experiment was performed to simulate construction tasks, which was conducted in a controlled indoor environment without major obstructions to avoid risk of interferences in the propagation of the wireless signal. Figure 2 shows a layout of the experimental testbed. In this experiment, one participant disassembles a deck and store material, another participant uses this material to assemble a raised deck in a different work area. Assembling and disassembling are dependent activities; two storage areas used by both participants and two installation areas used separately.

Two simulated construction tasks were performed in the same space using a similar experimental layout (see Figure 2). Four video cameras were installed on the perimeter of the experimental area for ground truth measurement. The participant's location and thoracic posture status are monitored using a UWB and a PSM system

The results of automatically detecting activities of the two participants are plotted in Figures 3. Two work zones (A2 and B2), three material zones (S1, S3, and S4), and four rest zones (R1 to R4) as well as the corresponding activities in and in between them were identified. The two participants worked as a team. While the first participant's duty was to de-install material from one work zone and deliver the material to the storage zones, the task of the second participant was to use the material available at the storage zones to install a floor system in another work zone.
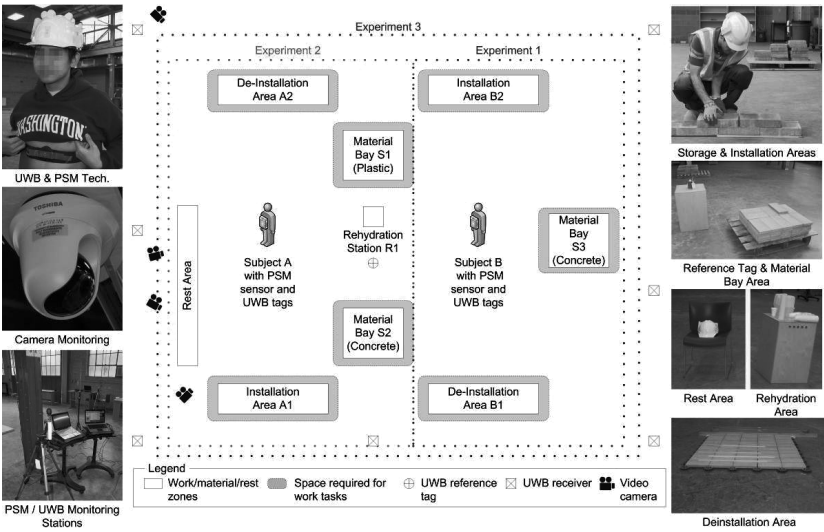
Figure 2: Experiment settings.
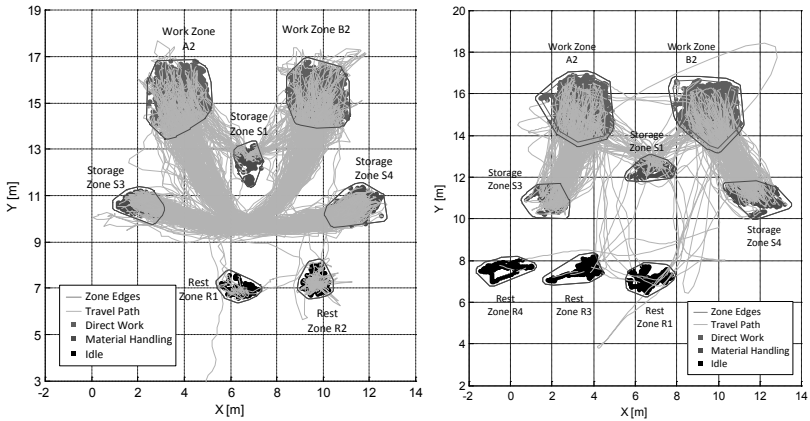
Figure 3: Work zones and trajectories of travel cycles of two participants.

Figures 4 and 5 show the direct work rate of both participants in this experiment. Participant 1 took four breaks during the experiment (see Figure 4). The total resting time was over 7 minutes in a work task that took about 90 minutes. More than 70% of the time was spent on traveling since the participant's duty was to deliver materials to the storage areas that the second participant used. The direct work time rate was therefore significantly smaller than in any of the two previous experiments. Manual study of video material and in particular measuring the times the first participant was traveling confirmed this observation. Figure 5 shows the direct work rate of Participant 2 in the same experiment. Since both participants were

conducting the activities at the same time, a correlation of the productivity performance can be noticed by the comparing the results with the direct work time rate. At the beginning of this experiment, Participant 2 (installing material) had to wait more than 40% of the first time segment for his team member (Participant 1) to set up the materials. Participant 1 (de-installing material) took two breaks during the 20-25 and 76-70 minute time segments since limited materials were available for de-installation (or in other words, Participant 1 completed the first de-installation task within approximately 22 minutes). Based on the information in Figures 4 and 5, Participant 2 had significantly more bending tasks to perform and took more frequently breaks. The reason is very likely the intense of the installation work that Participant 2 had to perform.
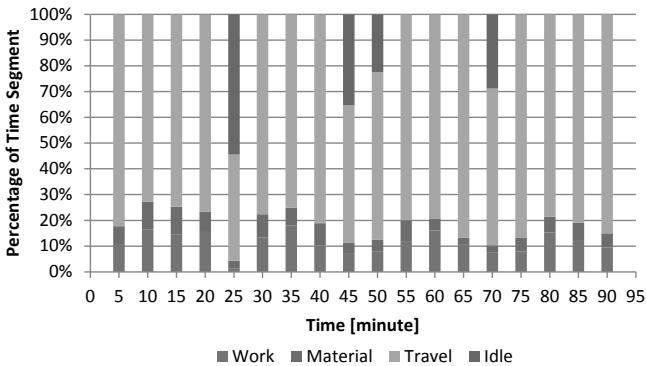
Figure 4: Result of automated work sampling, every 5 minutes (Participant 1).
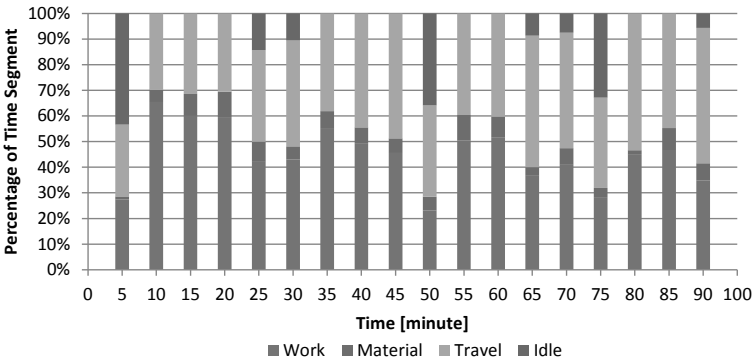
Figure 5: Result of automated work sampling, every 5 minutes (Participant 2).

CONCLUSIONS

Previous research has found that remote and rapid sensing such as Ultra Wideband (UWB) and Physiological Status Monitoring (PSM) technology can effectively facilitate automatic monitoring of the position, posture, and physiological

status of construction personnel. However, these technologies have not been used to improve productivity and activity assessment. The results of this paper show that current technology is satisfactorily reliable in autonomously and remotely monitoring participants during simulated construction activities. In addition, the authors have found that data from various sensing sources can be successfully fused to augment real-time knowledge of construction activity (and potentially productivity) assessment, which would reduce, if not avoid, the shortcomings of traditional visual observation and estimation of productivity rates. The output of the proposed approach could be used by contractors to evaluate the maximum actual production against the planned production as a way to automatize project control functions and perform true real-time "productivity and activity assessment". The real-time productivity and activity assessment will enable project managers to accurately determine the progress of construction operations and easily share the information with all project parties. In summary, the presented work has shown the potential of technologies lies in the integration of various technology-specific data sources. While technology manufactures are quickly improving the level of integration and the richness of data collected, research as the one described in this paper advances knowledge of data fusion for construction applications.

REFERENCES

Abeid, J. and Arditi, D. (2002). "Linking time-lapse digital photography and dynamic scheduling of construction operations." *Computing in Civil Eng.*, 16: 269-279.

Arditi, D. and Krishna, M. (2000). "Trends in productivity improvement in the US construction industry." *Construction Management and Economy*, 18:15–27.

Cheng, T. Migliaccio, G.C. Teizer, J. and Gatti, U. C. (2013). "Data Fusion of Real-time Location Sensing (RTLS) and Physiological Status Monitoring (PSM) for Ergonomics Analysis of Construction Workers." *Computing in Civil Engineering*, http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000222.

Cheng, T. Venugopal, M. Teizer, J. and Vela. P.A. (2011). "Performance Evaluation of Ultra Wideband Location Tracking Data in Harsh Construction Environment." *Automation in Construction*, 20(8), 1173-1184.

Construction Industry Institute (CII), Guide to activity analysis, University of Texas at Austin, Implementation Resource 252-2a. July 2010.

Goodrum, P.M. and Haas, C.T. (2002). "Partial factor productivity and equipment technology change at activity level in U.S. construction industry." *Construction Engineering and Management*, 128(6): 463-472.

Gouett, M. Haas, C. Goodrum, P. and Caldas, C. (2011). "Activity analysis for direct-work rate improvement in construction." *Construction Engineering and Management*. 137(2): 1117-1124.

Shahandashti, S.M. Razavi, S.N. Soibelman, L. Berges, M. Caldas, C.H. Brilakis, I. Teizer, J. Vela, P.A. Haas, C. Garrett, J. Akinci, B. and Zhu, Z. (2011). "Data-fusion approaches and applications for construction engineering." *Construction Engineering and Management*, 137(10): 863-869.

Teizer J. and Vela, P. A. (2009). "Personnel tracking on construction site using video cameras." *Advanced Engineering Informatics*, 23(4): 452-462.

High-precision and Infrastructure-independent Mobile Augmented Reality System for Context-Aware Construction and Facility Management Applications

Hyojoon Bae[1], Mani Golparvar-Fard[2] and Jules White[3]

[1]Bradley department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24060; email: hjbae@vt.edu
[2]Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801; email: mgolpar@illinois.edu
[3]Bradley department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, 24060; email: julesw@vt.edu

ABSTRACT

This paper presents a new context-aware mobile augmented reality system that provides rapid and robust on-site access to construction project information such as drawings, specifications, schedules, and budgets. The mobile augmented reality system does not need any RF-based location tracking (e.g., GPS or WLAN) or optical fiducial markers for tracking a user's position. Rather, the user's location and orientation are automatically and purely derived by comparing photographs from the user's phone to a 3D point cloud model created from a set of site photographs. After generating a 3D point cloud model of construction site, field personnel can use mobile devices to take pictures of building elements and be presented on-site with a detailed list of project information related to the visible construction elements in an augmented reality format. The experimental results show that (1) the underlying 3D reconstruction module of the system generates more complete 3D point cloud models, and faster than other state-of-the-art Structure-from-Motion(SfM) algorithms; (2) the localization method is an order of magnitude more accurate than the state-of-the-art solutions, and can provide acceptable tolerances of most on-site engineering applications. Using an actual construction case study, the perceived benefits and limitations of the proposed method for on-site context-aware applications are discussed in detail.

INTRODUCTION

On-demand and inexpensive access to project information has significant potential to improve decision-making during on-site construction activities. Despite the importance, current practice of monitoring on jobsites includes manual data collection, non-systematic analysis and visually/spatially complex reporting (Golparvar-Fard et al. 2012). Particularly during the data collection/analysis, field personnel have to carry large stacks of paperwork on jobsites and spent significant amount of time retrieving relevant project information. Such inefficiency in turn causes downtime/rework and can ultimately lead to schedule delay or cost overruns.

Over the past decade, several research projects have focused on providing project information to field personnel through mobile devices and/or augmented reality systems (Akula et al. 2011; Khoury & Kamat 2009). These works have primarily focused on using Global Positioning Systems (GPS), Wireless Local Area Networks (WLAN), Ultra Wide Band (UWB), or Indoor GPS for accurately positioning the user within congested construction environments. The main drawback of these RF-based location tracking technologies is their high degree of dependency on preinstalled infrastructure, which makes their application difficult for construction sites. Recently, some researches have focused on developing location tracking technologies that are infrastructure-independent. These infrastructure-independent systems are typically based on inertial measurements and make use of accelerometers and gyroscopes. Given their independence from an existing infrastructure, however, their application may result in accumulated drift error which grows with the distance traveled by the users (Akula et al. 2011).

Instead of using external location tracking components and/or installing new technological component to construction site, our proposed system uses only photographs of construction site to provide accurate location and orientation information of field personnel. The proposed vision-based system, Hybrid 4-Dimensional Augmented Reality (HD^{4}AR), is built on our previous work (Bae et al. 2012) which presents high-precision visualization of semantically-rich 3D cyber-information over real-world imagery in an augmented reality format as in Figure 1.

Rather than using imprecise mobile device's GPS and/or wireless sensors with tens of meters of accuracy, the HD^{4}AR system allows construction field personnel to use existing and already available camera-equipped mobile devices, such as smartphones or tablets, to take pictures for accurate localization. Compared to our previous work (Bae et al. 2012), this paper presents algorithmic enhancements that

(a) Sample Base Images   (b) 3D point cloud model   (c) Augmented Photographs
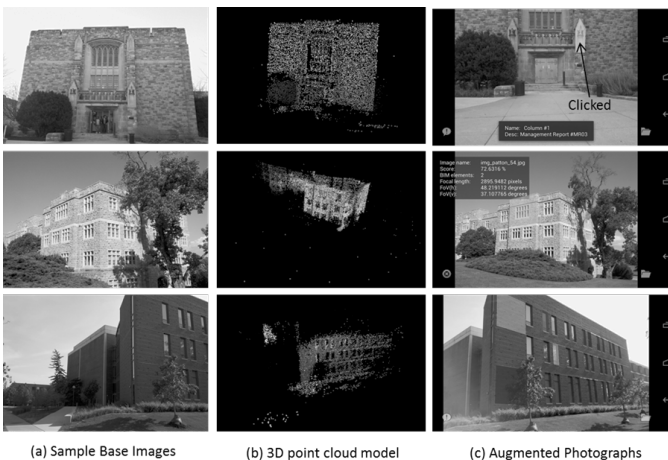
Figure 1. Examples of HD^{4}AR applications on different static buildings (adopted from Bae et al. 2012)

improve the speed of HD^{4}AR and the impact of image feature detection and description methods on localization success-ratio.

Given the popularity of smartphones and particularly their rising availability on construction sites, the HD^{4}AR, which relies on existing camera-equipped mobile devices, can be used to build many promising mobile augmented reality systems for context-aware Architecture/Engineering/Construction and Facility Management (ACE/FM) applications. The HD^{4}AR system does not require any precise location tracking modules but still can provide high-accurate localization results of mobile devices using photographs. This vision-based approach makes the system practical and inexpensive to apply in various construction sites.

The remainder of this paper is organized as follows: Section 2 presents an overview of the HD^{4}AR system. The details of the 3D reconstruction process that generates the 3D point cloud model from a set of photographs and the localization/augmentation process using a generated 3D point cloud are also discussed in Section 2. Section 3 presents empirical results from experiments with HD^{4}AR and also compares the performance to other state-of-the-art localization solutions. Finally, the perceived benefits and limitations are presented in Section 4. Video demos, detailed performance data, along with additional supplementary material for HD^{4}AR can be found at http://www.magnum.ece.vt.edu/index.php/research-projects/100-hd4ar and http://raamac.cee.illinois.edu/hd4ar.php.

HD^{4}AR: HYBRID 4-DIMENSIONAL AUGMENTED REALITY

Overview. In order to augment a given photograph with cyber-information, the HD^{4}AR requires that construction field personnel first take overlapping photos of the construction site to produce a physical model, i.e., 3D point cloud. The initial 3D point cloud generation is based on a Structure-from-Motion (SfM) algorithm that estimates the 3D position of the image features through feature extraction, matching, initial triangulation, and an optimization process called Bundle Adjustment.

After this initial bootstrapping is complete, the user can take a new photo at a random location and his/her location is determined by comparing this new image to the 3D point cloud. Once the user is localized, the fused cyber-physical model can be used to predict where cyber-information, such as elements of Building Information Model (BIM), should appear in the original photograph and where physical objects actually are in the cyber 3D space. This ability to track where the elements of BIM should appear in photographs serves as the foundation of the HD^{4}AR system. After a field engineer has an augmented photograph, he or she is able to use a multi-touch interface to select physical objects in view to retrieve more information associated to that object. Moreover, a field engineer can create new BIM elements by drawing a polygon on the photograph. Then the user-created 2D BIM elements are automatically converted to 3D elements (back-projection) and attached to the existing cyber-information. This 3D annotation functionality is also one of the distinct features of HD^{4}AR. Figure 2 summarizes the components of HD^{4}AR.

3D Reconstruction (Bootstrapping) Process. The 3D reconstruction procedure in HD^{4}AR mostly follows the original steps in SfM algorithm of the

Bundler package (Snavely et al. 2007) except that it 1) uses different feature detectors and descriptors, 2) introduces new optimization parameters for reducing noise in the 3D point cloud to improve localization accuracy, and 3) exploits multi-core CPU and GPU hardware for faster processing speeds. All the steps of 3D reconstruction in the HD[4]AR system are summarized as follows:

1) *Feature Extraction*: A feature detection and extraction algorithm is run on each base image to create a set of feature descriptors to be used as the basis for matching images to one another. The GPU-based Speeded-Up Robust Features (SURF) descriptor (Bay et al. 2008) and Fast REtinA Keypoint (FREAK) descriptor (Alahi et al. 2012) are implemented and used in the HD[4]AR.

2) *Feature Matching:* The set of feature descriptors extracted from each image is iteratively compared against each other. For initial matching, HD[4]AR creates a *k-d* tree of the descriptors and runs the Approximate Nearest Neighbors (ANN) searching algorithm (Arya & Mount 1998). Next, the HD[4]AR performs the ratio test as described in (Lowe 2004) to reduce the percentage of false matches. Finally, HD[4]AR robustly estimates a Fundamental matrix and removes outliers for every image pair using the RANSAC (RANdom SAmple Consensus) algorithm with the eight-point algorithm (Hartley & Zisserman 2004).

3) *Structure-from-Motion*: The SfM algorithm begins by estimating the camera parameters, such as focal length, rotation matrix and translation vector, of an initial image pair using Nistér's five-point algorithm (Nistér 2004), and the 3D positions of their feature points are triangulated. Then HD[4]AR system attempts to calibrate camera parameters of each additional base image from already triangulated 3D points and to register new points seen by calibrated new cameras. When the new image is registered, the Bundle Adjustment algorithm is run to minimize the overall re-projection error of all triangulated 3D points. Figure 3 shows an example of final 3D point cloud model resulting from the HD[4]AR 3D reconstruction process.
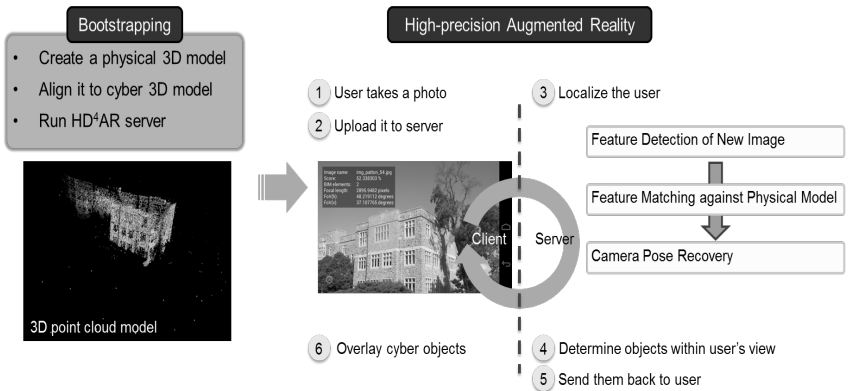
Figure 2. The overall procedures of the HD[4]AR system

Localization and Augmentation Process. Once the point cloud is generated, the system can augment photos sent from the HD[4]AR client running on a user's mobile device. From a high-level perspective, this process operates as shown in Figure 2. Step 1, the field personnel, upon finding a section of the worksite he/she wishes to query, takes a picture of the area using a mobile device. Step 2, the device uploads the captured image to the HD[4]AR server. Step 3, the server runs feature detection, feature matching, and camera calibration algorithms to identify the relative location of the camera against the 3D point cloud. Step 4, using the relative rotation and translation information of the image as input, the server determines what BIM elements are within the image's field of view, and where they appear. Step 5, the objects are sent back to the user's device with positional information and semantic information. Step 6, the user's device renders the captured image, overlaid with the returned objects in the correct locations for the device's location.

A key improvement over our previous work is that the HD[4]AR now finds the correspondences between image and 3D point cloud model by directly comparing feature descriptors of an image to that of each 3D point. This leads to significant improvements in localization speed. In our previous work, the server runs matching against an entire set of base images, which incurs unnecessary descriptor comparisons.

Figure 4 shows some example screenshots from the Android HD[4]AR client with Android ver. 4.2. In the screenshots the 3D BIM information is precisely aligned with the real-world imagery despite significant changes in the position and orientation of the user's device.

EXPERIMENTAL RESULTS

This section describes experiments we conducted to assess the ability of the HD[4]AR to produce initial 3D point clouds, localize new photographs taken at random locations, and displaying appropriate cyber-information on top of the photographs.

(a) Initial base images   (b) 3D point-cloud from HD[4]AR
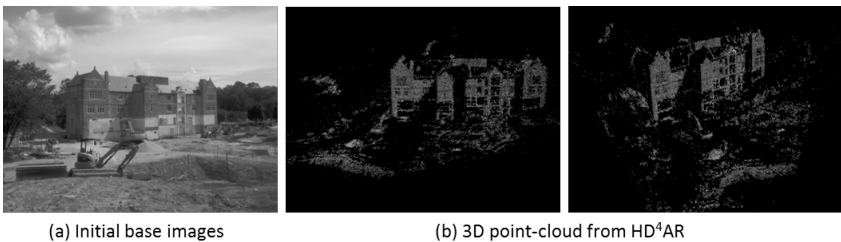
Figure 3. Example of 3D jobsite reconstruction result from the HD[4]AR.

Figure 4. Example of augmented photographs from the HD[4]AR.

Platform Specification and Data sets. The server side of the HD^{4}AR was running on a desktop computer with 8 gigabytes of 667 MHz DDR3 RAM, and a 4-core Intel i7 CPU 870 (@2.93 GHz) processor running Ubuntu version 12.04. The NVIDIA Geforce GTX 560 Ti graphic card was used for GPU computations. The image data sets used to create 3D point clouds came from the actual construction site, *Center for the Arts*, on campus of Virginia Tech. All the photographs were taken using Samsung Galaxy Nexus smartphone with Android version 4.2.

Performance of 3D Reconstruction. First, a 3D reconstruction procedure with HD^{4}AR was performed on data set. To show the computational speed of the HD^{4}AR, the same data sets were run on the Bundler package (Snavely et al. 2007), the most popular software package that implements SfM for 3D reconstruction. The overall computational time is compared in Table 1 and it shows that HD^{4}AR with FREAK descriptor is up to ~20 times faster than the Bundler. Figure 5 shows the result of 3D point cloud model.

Localization. The localization success-ratio and the speed using a 3D point cloud generated from the HD^{4}AR were also measured. The localization test images were taken at random locations and tested on-site for localization correctness. As observed in Table 2 and Figure 6, the result proves that the HD^{4}AR can successfully localize a user solely based on an image and within few seconds. Compare to the Bundler package, the HD^{4}AR with FREAK descriptor increases the speed of localization by an average factor of 29.41. Finally, we tested a 3D annotation functionality of the HD^{4}AR. As shown in Figure 7, a window drawn by user on 2D image is automatically back-projected to the 3D cyber space and correctly aligned with 3D point cloud. This back projection includes the feature extraction/matching of user-created polygons and triangulation of the polygon using camera information of registered base images. Once user-created information is successfully back-projected, it can be precisely overlaid on other photographs from different viewpoints as shown in rightmost screenshot.

Table 1. Elapsed time of 3D point cloud reconstruction

	Bundler package	HD^{4}AR (with SURF)	HD^{4}AR (with FREAK)
Number of registered images	125 / 125	125 / 125	125 / 125
Elapsed Time	182 *min*	12.49 *min*	9.25 *min*
Performance gain	×1.0	×14.58	×19.69
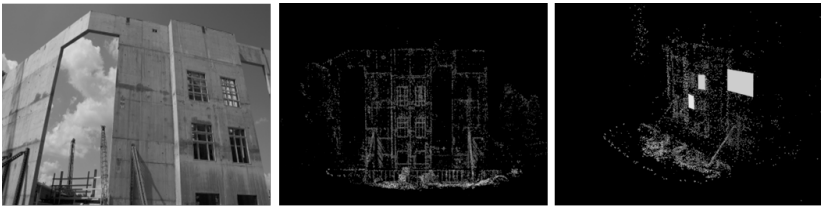
(a) Initial base images   (b) 3D point-cloud, HD^{4}AR (SURF)   (c) Aligned with cyber information

Figure 5. Resulting 3D point clouds from the HD^{4}AR.

Table 2. Localization success-ratio and average localization time

	Bundler package	HD^{4}AR (with SURF)	HD^{4}AR (with FREAK)
Localization success-ratio	100% (65/65)	98.46% (64/65)	95.38% (62/65)
Average Time	93.2 *sec*	3.47 *sec* (16.02 *sec*)*	3.17 *sec*
Performance gain	×1	×26.8	×29.41

* The result using our previous approach (Bae et al. 2012).

Figure 6. On-site localization results using Android smartphones. Cyber information is precisely overlaid in all different viewpoints.
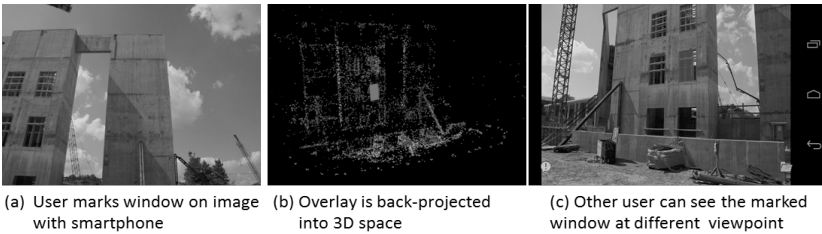
(a) User marks window on image with smartphone   (b) Overlay is back-projected into 3D space   (c) Other user can see the marked window at different viewpoint

Figure 7. Example of 3D annotation (daily construction reporting) in the HD^{4}AR.

CONCLUSION

The HD^{4}AR system was designed with the intent of bringing augmented reality to construction sites, to better facilitate the use of cyber-information around worksites. The HD^{4}AR allows field personnel to use existing mobile devices, such as smartphones or tablets, to take pictures for accurate localization. This in turn enables them to annotate pictures for daily construction reporting and store the information in 3D. It also enables them to query information such as specifications on demand by just taking a picture and retrieving augmented reality overlays. Given the popularity of smartphones and tablets and particularly their rising availability on construction sites, the vision-based and location tracking-free solution can lead to many promising mobile AR systems for context-aware AEC/FM applications. The HD^{4}AR works using image features as the basis for user localization, using SfM techniques to build and match a 3D geometric model from regular smartphone camera images. Users can use a smartphone outfitted with a camera, screen, and wireless communication to upload a captured image, localize it, and then overlay the returned cyber information on the physical objects in the photograph to which it pertains. The performance of the

HD^{4}AR, with a localization success-ratio of 98.46% (using SURF), implies that it is possible to develop a near real-time augmented reality systems using construction site photographs. It takes ~3 seconds for localization and less than 10 minutes for point cloud generation. With everyday data collection and application of the HD^{4}AR, 3D point clouds can be produced very quickly, allowing AEC/FM practitioners to easily monitor construction progress by quickly accessing relevant information. In future work, we plan to query full IFC-based (Industry Foundation Class) BIM rather than using manually created elements or daily construction reporting items.

ACKNOWLEDGEMENTS

REFERENCES

Akula, M. et al., (2011). "Integration of infrastructure based positioning systems and inertial navigation for ubiquitous context-aware engineering applications." *Advanced Engineering Informatics*, 25(4), 640–655.

Alahi, A., Ortiz, R. & Vandergheynst, P., (2012). "FREAK: Fast Retina Keypoint." 2012 *IEEE Conf. on Computer Vision and Pattern Recognition*, 510–517.

Arya, S. & Mount, D., (1998). "An optimal algorithm for approximate nearest neighbor searching fixed dimensions." *Journal of the ACM*, 1(212).

Bae, H., Golparvar-fard, M. & White, J., (2012). "Enhanced HD4AR (Hybrid 4-Dimensional Augmented Reality) for Ubiquitous Context-aware AEC/FM applications." *In 12th International Conference on Construction Applications of Virtual Reality (CONVR 2012)*. Taipei, Taiwan, 253–262.

Bay, H. et al., 2008. "Speeded-Up Robust Features (SURF)". *Computer Vision and Image Understanding*, 110(3), pp.346–359.

Hartley, R. & Zisserman, A., (2004). *Multiple view geometry in computer vision*, Cambridge University Press.

Khoury, H.M. & Kamat, V.R., (2009). "High-precision identification of contextual information in location-aware engineering applications". *Journal of Advanced Engineering Informatics*, 23(4), 483–496.

Lowe, D.G., (2004). "Distinctive Image Features from Scale-Invariant Keypoints." *International Journal of Computer Vision*, 60(2), 91–110.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012). "Automated model-based progress monitoring using unordered daily construction photographs and IFC as-planned models." *ASCE Journal of Computing in Civil Engineering*. (In press).

Nistér, D., (2004). "An efficient solution to the five-point relative pose problem." *IEEE Transactions pattern analysis and machine intelligence*, 26(6), 756–777.

Snavely, N., Seitz, S.M. & Szeliski, R., (2007). "Modeling the World from Internet Photo Collections." *International Journal of Computer Vision*, 80(2), 189–210.

Evaluation of the Position and Orientation of (Semi-) Passive RFID Tags for the Potential Application in Ground Worker Proximity Detection and Alert Devices in Safer Construction Equipment Operation

E. Marks[1] and J. Teizer[2]

[1] RAPIDS Construction Safety and Technology Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive N.W., Atlanta, GA 30332-0355; PH (859) 321-6542; email: ericmarks@gatech.edu

[2] RAPIDS Construction Safety and Technology Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive N.W., Atlanta, GA 30332-0355; PH (404) 894-8269; FAX (404) 894-2278; email: teizer@gatech.edu

ABSTRACT

Of the construction fatalities that occurred in 2010, approximately one-fourth resulted from workers being struck-by an object or piece of equipment on a construction site. The objective of this particular research is to evaluate the capabilities and reliability of low-cost (semi-) passive radio frequency identification (RFID) technology as the main component of a pro-activate real-time personal protection unit (PPU). Experiments emulating typical interactions between ground workers and heavy construction equipment are designed and executed. Various positions and orientations of the designed (semi-) passive RFID-based PPU were evaluated based on typical ground worker movements during construction tasks. Results indicate that both position and orientation of the PPU impact the reliability of the system's ability to activate alerts during hazardous proximity situations. The overall purpose of this research is to generate scientific data and knowledge of proximity detection and alert system for eventual implementation of these systems on construction sites. By implementing these systems in the construction environment, ground workers can be provided an alert during hazardous proximity situations.

INTRODUCTION

The dynamic nature of each construction site often requires multiple construction resources including personnel and heavy equipment to operate within close proximity to one another. Hazardous proximity conditions and situations exist when heavy construction equipment is operating in close proximity to ground workers. These conditions can potentially result in injuries and fatalities for ground workers caused by contact collisions between workers and heavy equipment. Previous research efforts have investigated interactions between heavy construction equipment and ground workers. Most of this work is focused on gathering and

analyzing injury and fatality statistics of contact collisions between ground workers and heavy equipment.

Both private constructions companies and government agencies have been slow in implementing automated technologies when compared to other industries. (Pratt et al. 2001). Other industries such as transportation, mining, and manufacturing have been actively implementing various prototype safety technologies. Construction safety research currently has a lack of scientific evaluation data and results gathered through experimental trials for safety technologies including proximity detection and alert systems (Teizer et al. 2010). Experimental trials designed to simulate ground worker movement on construction sites are required to adequately evaluate construction safety technologies and eventual deployment of these systems into the construction industry.

LITERATURE REVIEW

Construction sites often have limited workspace requiring construction resources to function in close proximity to one another. As construction injury and fatality statistics show, these work environment conditions result in dangerous proximity situations for ground workers. Proximity issues remain a key problem in the safety of workers on construction projects. The following review covers safety statistics of the construction industry, proximity detection and alert systems, current testing methods for these systems, and a research needs statement.

Construction Equipment and Human Interaction

When compared to other industries in the US, construction continues to rank among the highest for workplace fatalities per year. In 2011, the construction industry recorded 721 fatalities and 17% of these fatalities resulted from workers colliding with objects or equipment (CFOI 2011). Fatalities caused by collisions accounted for 2.6% of the nation's workplace fatalities in 2011. The construction industry averaged 197 fatalities per year resulting from workers being struck-by construction equipment or objects (CFOI 2009). Hazardous proximity situations between ground workers and construction equipment have been the focus of previous research efforts. The harsh outdoor environment and often repetitive nature of construction tasks causing workers to decrease their awareness have been identified as root causes for hazardous proximity situations (Pratt et al. 2001).

Proximity Detection and Alert Technology and Test Methods

Technologies thought to be capable of alerting construction personnel in real-time during hazardous proximity conditions include RADAR (Radio Detection and Ranging), Global Positioning System (GPS), sonar, vision, radio transceiver tags, etc. Each has a set of unique set of limitations when implemented on a construction site. These limitations include availability and signal strength, size, weight, and power source (Ruff 2005). Parameters such as alert method, precision, reliability, and alert distance were assessed for many proximity detection technologies when deployed in a simulated construction environment (Teizer et al. 2007). Active (battery-powered)

radio frequency technology displayed an ability to satisfy many of these evaluation parameters (Teizer et al. 2010).

Previous research efforts have also produced testing methods designed to evaluate the capabilities of proximity detection and alert systems. Ruff (2005) used manual methods to measure proximity alert distances for a camera and radar proximity detection system. These developed methods were implemented to evaluate proximity detection and alert systems deployed on large capacity haul trucks during a surface mining operation (Steel et al. 2003).

A lack of scientific evaluation data exists for automated safety technology for construction sites. A need was identified by Teizer et al. (2010) to evaluate combinations of mounting positions and orientations of proximity detection and alert system components. This evaluation should be accomplished through current or newly developed experimental methods, case studies, and data analysis.

OBJECTIVE AND SCOPE

The objective is to evaluate the reliability of various positions and orientations of passive RFID technology as a potential detection component in Personal Protective Unit (PPU) devices. As part of a proximity detection and alert system, its function is activating an alert in real-time to construction personnel when they get too close to an antenna deployed on a piece of construction equipment. The different positions and orientations of passive RFID for PPU were tested in simulated ground worker movements towards commercially-available components of a passive RFID tag-antenna system.

METHODOLOGY AND RESULTS

All experimental trials were designed to generate evaluation data detailing the effectiveness of various passive RFID tag positions and orientations for a PPU within a proximity detection and alert system. The PPU was mounted at various locations and orientations on Personal Protection Equipment (PPE).

Technology Evaluated

The real-time proximity detection and alert system evaluated during these experiments utilizes active Ultra-High radio frequency (UFH) technology to detect proximity breaches of construction equipment and ground workers. If two or more construction resources are in too close proximity, the sensing technology will activate an alarm to warn construction personnel through devices called Equipment and Personal Protection Units (EPU and PPU, respectfully). The EPU is equipped with a reader, alert mechanism, and single directional antenna that transmits and receives PPU information during a proximity breach such as a timestamp, PPU identification number and magnitude of the reflected radio frequency signal. The PPU is equipped with an alert mechanism, chip, and battery. This device can be installed on a worker's hard hat or safety vest. The PPU surface area dimensions are 5.5 cm by 8.5 cm and the thickness is minimal. The PPU is flexible and can be attached to solid objects. Both the EPU and PPU device scan be viewed in Figure 1.

Power is supplied to the EPU component of the proximity detection device through the existing battery on the piece of equipment. The PPU intercepts and reflects back a broadcasted signal from the EPU which instantaneously activates an alert in real-time when devices are in close proximity to one another. The audible alert triggered by the EPU creates ample noise so that the equipment operator is able to hear the alert above sounds typically to construction sites. Figure 1 shows the EPU installed on a tripod (left) and PPU mounted on a worker's safety vest (right).
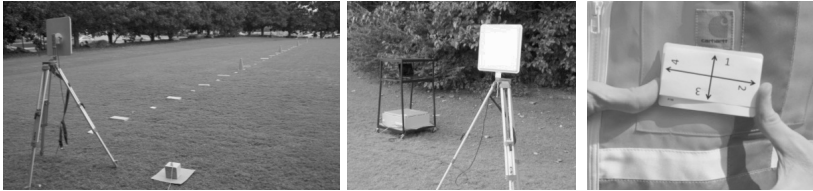
Figure 1. Proximity detection and alert system test bed: EPU and PPU.

The signal broadcasted by the EPU is projected in a radial manner and loses strength as the PPU moves farther from the EPU location. The signal strength emitted by the EPU remained consistent throughout the experimental trials. The EPU's antenna was installed in locations where the line-of-sight between the EPU and PPU was not obstructed. The proximity detection and alert system evaluated is also able to log data concerning PPU proximity breaches. The data logging function records the PPU and EPU identification number as well as the timestamp and Received Signal Strength Indication (RSSI) for each proximity breach.

Experiments

The objective of the designed and executed experiments was to identify the best position and orientation among multiple variations for the proximity detection and alert system tested. All experimental trials were performed in an outdoor environment with mostly clear and mostly sunny weather conditions with a temperature of 75 degrees Fahrenheit. A clear and flat grass ground surface with no obstructions was used for a test bed for these trials. A Robotic Total Station (RTS) was used to position markers along a straight path perpendicular to the face plane of the EPU antenna. As displayed in Figure 1, the markers (centerline) were placed at three meter intervals to outline the walking path of a test person towards the antenna. In later experiments a test person approached the antenna from angles within its field-of-signal (FOS), which was approximately 30 degrees to each side of the centerline.

The EPU's antenna component was mounted on a tripod with the face plane perpendicular to the ground surface. for the experimental trials. The EPU antenna is capable of reading 60 degrees in direction parallel to the antenna's face plane. The antenna's centriod was postioned 1.15 meters vertically from the ground surface. This vertical distance represented the average elevation between the top of test person's hard hat and center of the test person's safety vest. Although the mounting postion of the antenna can vary on equipment, testing with other heights was not performed.

The test person equipped with a semi-passive RFID tag for the PPU began walking outside of the proximity range (approximately 40 meters from the EPU) along the path outlined by the markers. The test person maintained a constant walking pace of approximately 3-4 m/s until an alert was activated. The test person stopped and measured the distance from the stopped position to the EPU's antenna position. Each combination of tag position and orientation was conducted ten times.

A total of eight PPU position and orientation combinations were tested. For the purposes of this research, the term "position" refers to the location of the face plane of the device in relation to the ground surface. For example, the horizontal position is achieved by the face plane of the device being parallel to the ground surface and perpendicual to the EPU antenna face. Likewise, a vertically positioned tag has the face plane perpendicular to the ground surface and parallel to the EPU antenna face. Figure 2 shows both of these tag positions where the device is mounted on a hard hat in the vertical position (left) and horizontal position (right).
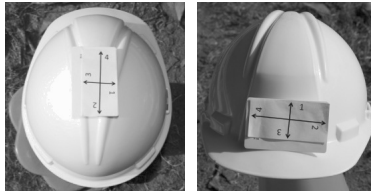
Figure 2. PPU mounted on a hard hat in the vertical position (left) and the horizontal position (right).

Four PPU orientations were used in combinations with the previously described PPU positions. The orientations were based on the location of the PPU in relation to the EPU antenna or the sky. Each of the four PPU oritenations were assigned a number (1, 2, 3, or 4) depending on the location of the PPU. Figure 3 presents a diagram to show how each numbered orientation was related to the EPU's antenna or sky and ground reference. Three different of the same type of tag were evaluated at the eight combinations of position and orientation.

Table 1 gives results of the statistical analysis of the alert distance results from the PPU positioned horizontally and mounted on top of the test person's hard hat. The value of the mean, minimum alert distance, range (which is the maximum value subtracted from the minimum value), and standard deviation of the alert disatnces were calculated from the ten trials of each PPU. The best performing values include the highest mean, and the lowest range and standard deviation (bolded in Table 1).
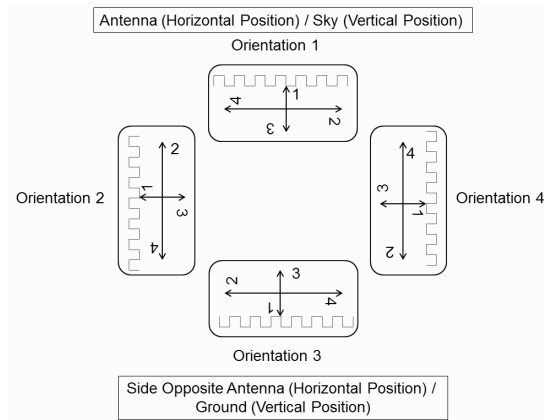
Figure 3. PPU orientation in relation to the EPU antenna and ground surface.

Table 1. Semi-passive tag Orientations mounted on top of a hard hat.

		PPU 1	PPU 2	PPU 3
Orientation 1	Mean:	**37.7 m**	**38.8 m**	**37.8 m**
	Min.:	36.8 m	38.0 m	37.0 m
	Range:	**1.3 m**	**1.5 m**	**1.3 m**
	Std. Dev.:	**0.6**	0.8	**0.7**
Orientation 2	Mean:	21.9 m	10.8 m	11.3 m
	Min.:	19.0 m	10.0 m	9.5 m
	Range:	5.8 m	**1.5 m**	3.0 m
	Std. Dev.:	2.9	**0.8**	1.6
Orientation 3	Mean:	34.1 m	12.7 m	37.4 m
	Min.:	32.0 m	12.0 m	36.3 m
	Range:	4.3 m	**1.5 m**	2.8 m
	Std. Dev.:	2.1	**0.8**	1.4
Orientation 4	Mean:	13.5 m	12.3 m	12.6 m
	Min.:	9.8 m	11.0 m	11.5 m
	Range:	5.5 m	2.8 m	5.3 m
	Std. Dev.:	3.0	1.4	2.7

Orientiation 1 or PPU 1 and 3 recorded the highest mean value, lowest range, and the lowest value for the standard deviation. Data was collected and analyzed for following eight individual configurations of vertical position: (1) side of hard hat, (2) front of hard hat, (3) back of hard hat, (4) front pocket of safety vest, (5) back pocket of safety vest, and (6) side of shoulder; and horizontally positioned, (7) top of worker hard hat, and (8) top of shoulder.

The best performing position was selected from the individual configurations tested. The best performers of each semi-passive RFID tag tested position and orientation configuration was identified from the four configurations tested on the

hard hat (see Table 2). A similar table was created to summarize the results from experimental trials completed using the safety vest. In Table 2, the top performers have bolded values, and the top performing orientations of each PPU is noted in parenthesis to the right of each value.

Table 2. PPU Orientation Summary on the Hard Hat.

		PPU 1	PPU 2	PPU 3
Top	Mean:	**37.7 m (1)**	**38.8 m (1)**	37.8 m (1)
	Min.:	11.5 m (4)	10.0 m (2)	12.5 m (2)
	Range:	**1.3 m (1)**	**1.5 m (1)**	**1.3 m (1)**
	Std. Dev.:	**0.6 (1)**	0.8 (1)	**0.7 (1)**
Side	Mean:	8.9 m (1)	5.7 m (1)	8.6 m (3)
	Min.:	1.5 m (3)	4.3 m (2)	4.5 m (2)
	Range:	1.0 m (1)	1.5 m (2)	5.3 m (2)
	Std. Dev.:	0.5 (1)	0.3 (3)	0.8 (2)
Front	Mean:	36.5 m (1)	38.2 m (1)	38.4 m (1)
	Min.:	4.5 m (4)	5.0 m (4)	12.5 m (4)
	Range:	3.0 m (1,2)	**0.5 m (3)**	1.5 m (2)
	Std. Dev.:	1.5 (1)	**0.3 (3)**	0.8 (2)
Back	Mean:	10.8 m (2)	28.8 m (3)	26.8 m (3)
	Min.:	2.0 m (4)	3.0 m (4)	4.0 m (2)
	Range:	1.3 m (2)	6.3 m (3)	7.8 m (3)
	Std. Dev.:	0.7 (2)	0.3 (4)	1.0 (1)

Orientation 1 mounted on top of the hard hat recorded the highest mean value when compared to the other tested configurations. When mounted on the side and front of the hard hat, orientation 3 experienced the lowest value of standard deviation. When mounted on top of the hard hat, the tag had the highest number of top performing orientations when compared to the other configurations tested. The front mounting position had similar values to mounting the tag on top of the hard hat. The largest mean value was recorded when placing the tag in the front pocket of the vest.

Further data anlysis showed that tag orientations 1 and 3 on the top and front of the hard hat were the best performers when compared to the other configurations evaluated. When mounted on the safety vest, orientations 1 and 3 had the highest alert distance. Other trials of mounting the PPU on the shoulder and back of the safety vest resulted in false negative alerts meaning the test person was able to reach the EPU antenna before an alert was activated. Polarization effects were experienced when the PPU was positioned vertically in orientation 2 and 4, locations that could prove benefical for other worker movements such as bending. In brief, the audible alert of the system was to a sufficient volume as to be heard over other noise typical of construction sites. In summary, tag position and orientation of (semi-) passive RFID tags played a key role in succesfully detecting the tag by the reader. The results demonstrate that deploying a single (or multiple) tag(s) may result in false or no reads.

More testing is required for these tag locations such as a test person approaching the EPU antenna while positioning his/her body such that their back is

facing the EPU antenna. Multiple tags can also be mounted on the worker hard hat and safety vest to better cover all approach angles. Further research is also required to evaluate potential external influences on the system such as the EPU antenna mounting location, impact of other construction resources, and various calibrated alert ranges. Data that is recorded and analyzed from these systems can improve safety site layout and assist in the development of new safety concepts for worker safety training.

CONCLUSION

Collsions between ground workers and construction equipment or objects is one of the leading causes of fatalities in construction. The purpose of this research was to evaluate the effectiveness and capabilities of various configurations of (semi-) passive RFID tags used for electronic personal protective device postions and orientations. Results from the review and experiments indicate that proximity detection and alert systems that systems relying on passive RFID position and orientation impact the effectivness of these systems. The major shortcoming that was identified thorugh the field trials is that commercially existing (semi-) passive RFID tags – based on its two-dimensional antenna design – may **not** be suitable for reliable implementation in PPU safety systems unless all potential worker-to-equipment poses and positions can be covered. Although existing construction research on active and passive RFID technology has shown mostly the benefits of RFID technology, further study is necessary, in particular on three-dimensional (semi-) passive RFID tags that avoid the previsouly described limitations of 2D RFID antenna tag designs.

REFERENCES

Census of Fatal Occupational Injuries (2011). "Fatal occupational injuries by industry and selected event or exposure." Bureau of Labor Statistics, <http://www.bls.gov/iff/oshcfoil.htm> (Oct. 21, 2012).

Census of Fatal Occupational Injuries (2009). "Fatal occupational injuries by industry and selected event or exposure." Bureau of Labor Statistics, <http://www.bls.gov/iff/oshcfoil.htm> (Oct. 25, 2012).

Pratt, S., Fosbroke, D., and Marsh, S.M. (2001). "Building Safer Highway Work Zones: Measures to Prevent Worker Injuries from Vehicles and Equipment." *Department of Health and Human Services,* 1(1): 5-6.

Ruff, T. (2005). "Evaluation of a Radar-Based Proximity Warning System for Off-highway Dump Trucks." *Accident Analysis and Prevention*, 38(1): 92-98.

Steel, J., Debrunner, C., and Whitehorn, M. (2003). "Stereo Images for Object Detection in Surface Mine Safety Applications." *Western Mining Resource Center,* Tech Report # TR20030109, Colorado School of Mines, Golden, CO.

Teizer J., Caldas C., Haas C. (2007) "Real-Time Three-Dimensional Occupancy Grid Modeling for the Detection and Tracking of Construction Resources." ASCE Journal of *Construction Engineering and Management*, 133: 880-888.

Teizer, J., Allread, B.S., Fullerton, C.E., and Hinze, J. (2010). "Autonomous Pro-Active Real-time Construction Worker and Equipment Operator Proximity Safety Alert System", *Automation in Construction*, 19(5): 630-640.

Development of a Navigational Algorithm in BIM for Effective Utility Maintenance Management of Facilities Equipped with Passive RFID

A. Costin[1], A. Shaak[2], and J. Teizer[3]

[1] Ph.D. Student, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N. W., Atlanta, GA, 30332; email: aaron.costin@gatech.edu
[2] Undergraduate Student, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N. W., Atlanta, GA, 30332; email: ashaak@gatech.edu
[3] Associate Professor, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. N. W., Atlanta, GA, 30332; PH (404)-894-8269; email: teizer@gatech.edu

ABSTRACT

Navigation through unfamiliar facilities, with labyrinths of corridors and rooms, is difficult and often results in a person being lost. The proposed research is to develop a navigation algorithm to help a person navigate through facilities. The scope is the use of passive radio frequency identification (RFID) tracking technology and Building Information Modeling (BIM). A novel theoretical approach is presented that utilizes current localization techniques, performance characteristics of passive RFID, and a BIM model. A prototype application has been developed that connects the RFID readers with a BIM model. The goals are to locate and navigate any user and/or utility in an unfamiliar facility. Preliminary results demonstrate the feasibility of locating a user inside buildings in real-time. Significantly, the algorithm has the potential to enable anybody to navigate the shortest and quickest route, saving time, money and, in an event of an emergency, lives.

PROBLEM DESCRIPTION

The goal of any facility is aimed at optimizing the client and end-user operation and management, and it is shown that technology and innovation improves functionality, sustainability and flexibility of facility components (Olatunji and Sher 2009). Additionally, building systems are becoming increasingly complex causing challenges for the management and operation of the facility (Kean 2011). Maintenance is an on-going process that runs throughout the project life cycle and requires continuous navigation throughout the facility. However, for a person unfamiliar with a facility, such as a maintenance manager or sub-contractor, locating a specific utility within a facility is often a tough task that leads to loss time and productivity. Also, it is of the utmost importance that the facility can be navigated quickly in the event of an emergency, because search and rescue have no time to

waste in getting lost when human lives are in need of rescue or evacuation. Current approaches for localization and navigation rely heavily on the facility's infrastructure, but if the facility infrastructure were to fail (i.e. power outage), the systems would fail. Therefore, there is a need for a location and navigation system that is self-dependent.

BACKGROUND

Building Information Modeling (BIM). Integrated building technologies allow a convergence and integration of systems to play a greater role in overall building performance (Kean 2011). BIM integrates the geometric and parametric properties of the 3D model of a building with all the information and properties of that building, such as product information, site schedule sequencing, and owner histories (Eastman et al 2008). The use of BIM has increased dramatically over the years, in which 71% of construction companies, 70% of architects, and (74%) contractors are using BIM, and is expected to keep increasing (McGraw-Hill 2012).

Radio Frequency Indoor Localization and Identification (RFID, UWB, and WLAN). RFID, UWB, and WLAN technology has been shown to be successful in a variety of applications including asset tracking, inventory management, on-site security upgrades, and productivity analysis (Goodrum et al 2006; Rueppel and Stuebbe, 2008; Li and Becerik-Gerber, 2011; Cheng et al. 2011; Taneja et al. 2011). The passive (without battery) RFID, with lifetime longevity and inexpensive costs, is most suitable for an indoor application of tracking workers or materials (Costin et al. 2012a). RFID technology has been successfully linked to BIM for supply chain management Sawyer (2008). There are other technologies that are being researched, but require the additional resources provided by the facility (Costin et al 2012b). Passive RFID does not require facility resources and can still perform in power outages. The investigation of a new method using passive RFID is required to transform the field of (construction) safety assessment, training, and education.

Indoor Localization. Haehnel et al. (2004) integrate indoor tag locations and laser scan data to produce accurate maps of RFID tags, which can be used for accurate localization of the robot and moving objects without odometry information. Current techniques to determine the location of a passive RFID tag require additional labor-intensive enhancements, such as sensor histogram models or pre mapping of tags (Costin et al 2012b). Therefore, to eliminate the need for these additional enhancements, Deyle et al. (2008) research developed a system that utilizes a robot that houses a physical sensor model and a particle filter. Moghtadaiee et al. (2011) uses FM radio signals. However FM signals do not carry any timing information, which is a critical factor in range calculation. Bouet et al. (2008) surveys the current state-of-art of RFID localization techniques and concludes that the choice of technique and RFID technology significantly affects not only the granularity and accuracy of the location information, but also the whole cost and the efficiency of the RFID system. Therefore, the best suited technique depends on the final application.

LOCALIZATION ALGORITHMS

Friis Transmission Equation. The Friis transmission equation (Ahson et al. 2008) is used to calculate the power received by an antenna, when transmitted from another antenna. Tracing the power to and from the reader can provide insights into both tag detection and received signal strength indicator (RSSI). The key elements involving physical aspects of antennas are antenna gains, reflection coefficients, and polarization. However, in real world applications, other factors include the effects of impedance mismatch, misalignment of the antenna pointing, multipath, atmospheric conditions, absorption, shadowing, diffraction, and material properties.

SLAM. Simultaneous localization and mapping (SLAM) is a popular technique used for autonomous vehicles to use sensing technology to build a map within an unknown environment (without *a priori* knowledge) while simultaneously using the map to compute its location. There are various forms of SLAM implemented in different environments, but the basic formula and structure are the same. The basic parts of SLAM include landmark extraction, data association, state estimation, state update and landmark update. Durrant-Whyte and Bailey (2006a) use the recursive Bayesian formulation for the real-time probabilistic estimation of location at a time instant k.

Received Signal Strength (RSS). Received signal strength (RSS) methods determine the location of the tag based on the signal strength received by the antennas. The signal strength propagates and reduces when the distance between the antenna and tag is increased. For instance, the further away a tag is, the lower the RSS. The maximum RSS is the power output from the reader, and the minimum is the signal strength needed to operate the tag. Therefore, knowing the output power from the antenna and the performance characteristics of the antennas and tags, an approximation of the tag location can be made. However, there are many factors and obstructions that affect the RSS indicators (RSSI), causing errors in the estimations (Fink and Beikirch 2011, Wang 2011). Therefore, additional information is needed to detect the characteristics of the RSS, such as computing RSSI "radio maps" or "probability maps" (Wang 2011). For accurate location detections, every indoor positioning system needs an underlying map for reference (Schafer et al. 2011).

METHODOLOGY

Framework and Apparatus. The scope is the use of passive radio frequency identification (RFID) tracking technology and Building Information Modeling (BIM). A prototype application has been developed that connects the RFID readers with a BIM model and a Microsoft Access database. The RFID-BIM framework has successfully demonstrated the feasibility of locating a user inside buildings in real-time (Costin et al. 2012b). The apparatus that houses the RFID equipment is a mobile cart that mimics one of a facilities manager (see Figure 1). Unlike a robot, the end user would push it throughout the facility. The cart comprises of one ThingMagic M6 UHF RFID reader that connects to four MTI MT-262006/TRH/A 902-928 MHz circular polarized antennas, a wireless router, battery, and laptop computer. The

design of the cart and placement of the antennas were determined based on experimental data to get the greatest view of read range.
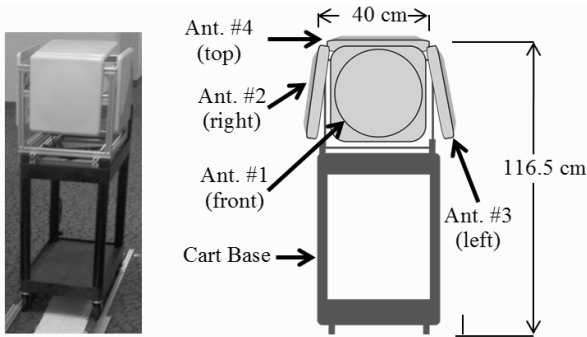
Figure 1. Mobile cart (left) and front view with dimensions (right).

Navigation Algorithm. In order to navigate a facility, the current position at time k must be determined. Therefore, a navigation algorithm is an extension of a localization algorithm. Unlike probabilistic SLAM that does not have *a-priori* knowledge, this approach utilizes the available BIM model in which the landmarks (passive tags) will be known. Having the exact tag locations gives the advantage of not needing to estimate the true location of the tags, drastically reducing the complexity of the algorithm. Therefore, the SLAM recursive time-update algorithm and measurement update algorithm can be derived, respectively:

$$P(x_k|Z_{0:k-1}, U_{0:k}, x_0, m)$$
$$= \int P(x_k|x_{k-1}, u_k) \times P(x_{k-1}|Z_{0:k-1}, U_{0:k-1}, x_0, m)\partial x_{k-1} \tag{1}$$

$$P(x_k, |Z_{0:k}, U_{0:k}, x_0, m) = \frac{P(z_k|x_k, m) \times P(x_k|Z_{0:k-1}, U_{0:k-1}, x_0, m)}{P(z_k|Z_{0:k-1}, U_{0:k})} \tag{2}$$

$X_{0:k}$ is the set of the vehicle location histories, $U_{0:k}$ is the set if the of control input histories, m is the set of all landmarks, $Z_{0:k}$ and is the set of all landmark observations. x_k is the state vector describing the location and orientation of the vehicle. u_k is the control vector, applied at time $k-1$ to drive the vehicle to a state x_k at time k. m_i is a vector describing the location of the ith landmark whose true location is assumed time invariant. z_{ik} is an observation taken from the vehicle of the location of the ith landmark at time k. When there are multiple landmark observations at any one time or when the specific landmark is not relevant, the observation will be written simply as z_k.

Passive RFID tags do not transmit power, but rather reflect the power back to the antenna, called incident power P_{tag}^{inc}. In order for a passive tag to operate, the reader must supply the operating energy to the tag through the transmission of an RF wave. In order to account for additional losses and multipath, the Friis transmission equation can then be derived to the Friis forward-link equation in Deyle et al. (2008).

$$P_{tag}^{inc} = P_{rdr} \cdot CL \cdot G_{tag} \cdot \left[\sum_{all\ paths} G_r(\theta) \cdot \left(\frac{3 \cdot 10^2}{4\pi \cdot d \cdot f} \right)^2 \right] \qquad (3)$$

P_{rdr} is the power transmitted by the antenna, which can be adjusted by the user. G_{tag} is the gain of the tag, which is specified by the manufacture. CL is the total loss from the cables, which can be calculated by the manufacture's specification. f is the operation frequency and d is the distance measured along the RF propagation path. The gain $G_r(\theta)$ is a polar function of elevation angle θ and azimuth angle ϕ, and can be determined by the two angles of the point relative to the antenna.

The RSSI, alone, cannot pinpoint the location. Therefore, the accumulation of RSSI's creates a 3D probability map of tag locations. Probability maps, or radio maps, are mathematical maps of the RSSI's for various distances and angles relative to the antenna. Therefore, a RSSI value by the reader can be matched to the maps, giving the possible locations of the tag. Moreover, each antenna has its own map, and since the view angles (beam width) overlap, the probability maps will also overlap (see Figure 2). These maps used with the Friis forward-link equation (3) and the known power output result in the needed information to estimate the tag location.
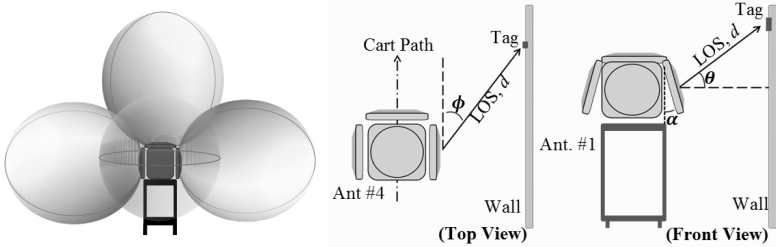
Figure 2. 3D RSSI probability map (visualization) showing overlap from each antenna (left). Line of sight (LOS) of the antennas to a tag showing the azimuth angle ϕ (middle) and elevation angle θ (right).

The algorithm first generates a list of all the tag locations from the BIM model. This list is the set of all landmarks, m, which will be continuously referenced. At current time k, the reader sends a signal and receives a list of tags that were read at that time. The list includes tagID, RSSI, frequency, and antenna of read, which are all used for probability estimations. This information can be passed through a particle filter (estimation algorithm) to give the highest probability location of that tag at current time k. Estimations include tag location in reference to the cart and the LOS distance, d. Additionally, position at time $k - 1$ will be used to estimate position at time k. These estimations are then referenced with the list of landmarks, and the best location is chosen, giving a (X,Y,Z) coordinate of the cart. This coordinate is then placed into the BIM model, where the location can be seen in real-time.

PRELIMINARY TESTING AND RESULTS

The first step is to measure and determine the performances of the passive RFID systems. Most RSS methods operate the reader at a high output power. Using the highest allowable power results in receiving a large number of tag reads, large amounts of noise, and backscatter. Reducing the power output from the transmitter limits the range of the signal, ultimately reducing the amount of backscatter and noise and improving the accuracy of the radio maps. In this approach, we limited the power output, P_{rdr}, to 27dBm. This level was previous determined by experimental results to include a minimal read range and an adequate number of tags with limited backscatter. The maximum read range at 27dBm is about 420cm, which will be emitted radially from the cart with all four antennas.

The system was first configured to determine the RSSI probability maps. Tags were set up on the walls, and reads were recorded continuously as the cart was pushed along a track. Various experiments were conducting, testing different variables, such as angles, distances, and power levels. The upper bound limits of the probability maps, measurements were taken in the most favorable conditions (e.g perfect pose, direct LOS, no interference etc.). The max RSSI was recorded to be -36 dBm at 10cm from the reader (see Table 1). The max beam width angle for the antennas is 65 degrees. Figure 3 shows how the RSSI decreases with the increase of line of sight LOS distance, d. However, having a low RSSI does not necessarily mean the tag is far away. Other factors, such as tag pose or interfere, can result in a low RSSI. For instance, a RSSI of -41 will result in a distance of around 100cm, but a RSSI of -75 can lie within 100-200cm. Therefore, the higher the RSSI, the higher the probability that the tag is close to the reader.
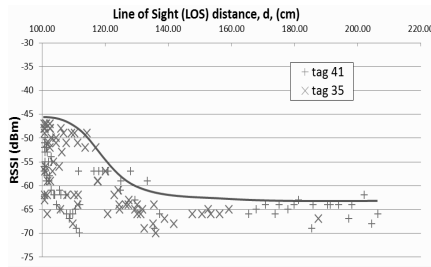
Figure 3. RSSI respect with LOS distance

Table 1. RSSI (dBm) in Most Favorable Conditions

LOS (cm)	Max.	Min.	Avg.	Stdev
10	-36	-40	-37.13	0.91
50	-45	-49	-46.66	1.42
100	-46	-50	-48.00	1.31
150	-53	-57	-54.72	1.40
200	-57	-61	-58.43	1.16
250	-57	-61	-58.17	1.30
300	-60	-66	-63.05	1.86
350	-64	-67	-65.53	1.07
400	-65	-67	-66.26	0.49

The system was deployed in an open hallway (2.7m high by 2.4m wide), to test the updating SLAM equations (1) and (2) using the RSSI. The corridor was also modeled in BIM. Tags were placed at a height of 101cm (keeping θ constant). The cart was positioned in direct center of the corridor and kept a straight line. The maximum error of the estimated coordinates was at most ½ the distance between adjacent tags (roughly 1 meter). Figure 4 shows the approximate location of the cart in the model.
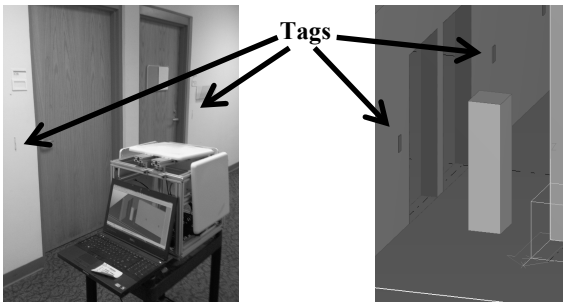
Figure 4. Actual location (left) and shown in BIM (right).

CONCLUSION

This paper presents a novel methodology for the development of a navigational algorithm that utilizes current indoor positioning and localization techniques with a BIM model for facility management and maintenance. Preliminary tests confirm the feasibility of the proposed approach and also that the general location of the cart was updated in real time in the BIM model. However, the characteristics of passive RFID present challenges for pinpointing the true locations, such as miss reads or lower RSSI. Fortunately, the complex probability equations in the algorithm estimate the location within reasonable error. Further refining is required to correct these challenges, including adjustments to the RSSI maps and the Friis forward-link equation.

The limitation of this research is that it assumes that there exists a BIM model of the facility. Future research is needed to create an algorithm that does not rely on known tag positions, for instance if the tags are not presented in the BIM model. Therefore, the algorithm can still locate the user in the building. Additionally, this would lead to research that automatically maps the tag positions into the BIM model.

REFERENCES

Ahson, S., and Ilyas, M. (2008). *RFID Handbook: Applications, Technology, Security, and Privacy.* CRC Press, Boca Raton.

Bouet, M., and dos Santos, A. L. (2008).RFID tags: Positioning principles and localization techniques. IEEE 1st IFIP Wireless Days. UAE, p. 1-5.

Cheng, T., Venugopal, M., Teizer, J., and Vela, P.A. (2011). "Performance Evaluation of Ultra Wideband Technology for Construction Resource Location Tracking in Harsh Environments", Automation in Construction, Elsevier, 20(8), pp. 1173-1184.

Costin, A., Pradhananga, N., Teizer, J. (2012a). "Leveraging passive RFID technology for construction resource field mobility and status monitoring in a high-rise renovation project." Automation in Construction, Elsevier, 24, 1-15.

Costin, A., Pradhananga, N., and Teizer, J. (2012b). "Integration of Passive RFID Location Tracking in Building Information Models (BIM)." EG-ICE, Int. Workshop, Herrsching, Germany, 4-6.

Deyle, T., Kemp,C., and Reynolds, M.S. (2008). "Probabilistic UHF RFID tag pose estimation with multiple antennas and a multipath RF propagation model." IEEE Int. Con. on IRS, Nice, France, Sept., 22-26, p. 1379-1384.

Durrant-Whyte, H., and Bailey, T. (2006a) "Simultaneous localization and mapping: part I," *Robotics & Automation Magazine, IEEE* , 13(2), p. 99-110.

Eastman, C., Teicholz, P., Sacks, R. and Liston, K., (2008). *BIM Handbook – A Guide to Building Information Modeling*. John Wiley & Sons, Inc., N.J.

Fink, A., and Beikirch, H. (2011) "Analysis of RSS-based location estimation techniques in fading environments," Indoor Positioning and Indoor Navigation (IPIN), Sept 21-23, p. 1-6.

Goodrum, P. McLaren, and M. Durfee. A. (2006). "The Application of Active Radio Frequency Identification Technology for Tool Tracking on Construction Job Sites." Automation in Construction, 15, p. 292-302.

Haehnel, D., Burgard, W., Fox, D., Fishkin, K. K., and Philipose, M. (2004). "Mapping and Localization with RFID Technology." IEEE International Conference on Robotics and Automation, p. 1015-1020

Kean, T. (2011). "Current Trends and Future Outlook Examining the IFMA industry trend report." Military Engineer, 103(673), p. 67-68.

Li, N. and Becerik-Gerber, B. (2011). "Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment", Advanced Engineering Informatics, 25(3), p. 535-546.

McGraw-Hill (2012). "SmartMarket Report: The Business Value of BIM in North America Multi-Year Trend Analysis and User Ratings (2007–2012) Construction." McGraw-Hill Construction.

Moghtadaiee, V., Dempster, A.G., and Lim, S. (2011). "Indoor localization using FM radio signals: A fingerprinting approach." Indoor Positioning and Indoor Navigation (IPIN), Sept. 21-23, p. 1-7.

Olatunji, O. A., and Sher, W. (2009). "Process problems in facilities management: an analysis of feasibility and management indices." BuHu 9th International Postgraduate Research Conference, Greater Manchester, January, 199-211.

Sawyer, T. (2008). "$1-Billion Jigsaw Puzzle Has Builder Modeling Supply Chains." http://enr.construction.com/features/technologyEconst/archives/080423-1.asp.

Rueppel, U. and Stuebbe, K.M. (2008). "BIM-Based Indoor Emergency-Navigation-System for Complex Buildings." Journal of Tsinghua Science and Technology, 13(S1), p. 362-367.

Schafer, M., Knapp, C., and Chakraborty, S. (2011). "Automatic generation of topological indoor maps for real-time map-based localization and tracking," Indoor Positioning and Indoor Navigation (IPIN), Sept. 21-23, p. 1-8.

Taneja, S., Akcamete, A., Akinci, B., Garrett, J.H., Soibelman, L., East, W.E. (2011). "Analysis of Three Indoor Localization Technologies for Supporting Operations and Maintenance Field Tasks", Computing in Civil Engineering, 26, p. 708-719.

Wang, H. (2011). "Bayesian radio map learning for robust indoor positioning." Indoor Positioning and Indoor Navigation (IPIN), Sept. 21-23, p. 1-6.

Identification of Construction Cumulative Trauma Disorders: A Machine Vision Approach

Fei Dai[1] and Xiaopeng Ning[2]

[1]Dept. of Civil and Environmental Engineering, West Virginia University, P.O. Box 6103, Morgantown, WV 26506; email: fei.dai@mail.wvu.edu
[2]Dept. of Industrial and Management Systems Engineering, West Virginia University, P.O. Box 6070, Morgantown, WV 26506; email: xiaopeng.ning@mail.wvu.edu

ABSTRACT

Cumulative trauma disorders (CTDs) are a leading cause of disability which generates huge direct and indirect costs in the construction industry. Existing efforts identify the risk of CTDs through survey or observatory studies which are subjective and expensive. This paper attempts to automatically identify the risk of CTDs on construction sites with the aid of a video surveillance system. Specific emphasis is placed on low back disorders (LBDs) which is one of the major work-related CTDs. To this end, a risk identification method is proposed based on location changes of region of interests in the resulting video streams at minimum supervision. It is used to detect repetitive trunk flexion-extension motions, awkward trunk postures, and assess risk of job-related LBDs. Preliminary experimental analysis on lab datasets demonstrates the promise of the proposed method.

INTRODUCTION

Cumulative trauma disorders (CTDs) known as work-related musculoskeletal disorders, repetitive strain injuries, repetitive motion disorders, and overuse syndrome are one of the largest causes of occupational disease in the United States (BLS 2012). In the construction industry, the high physical demands of strength and endurance expose construction workers to many of recognized risk factors such as repetitive motions, high force exertions, and awkward body postures (Kisner and Fosbroke 1994), leading to a high incidence rate (16 cases per 1000) of CTDs among the construction workers (BLS 2012). Note that the above statistics do not include those unreported and cases that do not cause loss of workdays. A recent study (Boschman et al. 2012) indicated that the incidence rate of CTDs in the construction industry can be as high as 67%, further reflecting the severity of the CTDs in construction.

Current research identifies the risk factors of CTDs primarily through surveys or observatory studies to understand the site workers' behaviors that induce CTDs. However, the methods in these efforts are subjective (either self or expert) assessment and therefore cannot be performed frequently and continuously. On the other hand, at construction sites, video surveillance systems already exist and have been routinely

used to monitor a project progress and oversee activities of site workers. To address the limitations of current methods and maximally utilize the existing resources at site, this paper makes an attempt to identify the risk of CTDs on construction sites with the aid of a single video surveillance and a biomechanical model that is developed by the authors. Specific emphasis is placed on the identification of risks for low back disorder (LBD) which is one of the major work-related CTDs.

RELATED WORK

Thus far, efforts on construction CTDs risk management have been primarily focused on: 1) promoting workplace ergonomic practice (e.g., OHBWC 2012; NCDL 2012), and 2) providing safety trainings (e.g., NIOSH 2007; OSHA 2012). These efforts are useful but have received limited effects. Current construction projects entail a sheer volume of manual, tedious and repetitive processes, giving rise to existing safety management efforts unable to effectively control and eliminate the potential risk. As a result, investments in ergonomic practice and safety trainings have been increasing annually, but the incidence rate regarding CTDs among construction workers has not seen decrease accordingly (BLS 2011; 2012). Besides, the industry and authorities have attached more importance to fatal accidents such as falls, cut, and burning. Very few measures have been taken for the construction daily activities to ensure rigorous assessment of ergonomic risks and therefore to prevent possible occurrence of CTDs (Inyang et al. 2012).

To alleviate this situation, research has been undertaken in assessing the risk of CTD injury. In the past three decades, a number of assessment models have been developed to evaluate joint and tissue loading. Most of these models require sophisticated instrument and laboratory environment in order to collect body motion and muscle activity data (Kingma et al., 1996; Marras and Granata 1997). Despite having high accuracy, the applications of these techniques in real work settings are rare. In order to assess the CTD risk on site, observational tools were developed (McAtamney and Corlett 1993; Hignett and McAtamney 2000). However, all observational tools require experts' visit and subjective evaluation. Thus, assessment cannot be performed continuously and only a limited number of jobs can be assessed during experts' visit. In addition, the inter-rater difference may result in disagreement in the results of their evaluations. More recently, video-based biomechanics tools were developed (Chang et al., 2003; Coenen et al., 2011). These new tools use single camera to acquire the human movement footage and estimate joint loadings base on whole body kinematics. However, these video-based tools are not considered fully automatic because they still rely on manual operation in the determination of posture and joint angles therefore cannot generate real-time feedback. Moreover, only a few key frames are identified for each lifting task which translates to a very low frame rate. As a result, it is very difficult to estimate the instantaneous load acceleration which has critical impact on joint loading (Marras and Granata 1997).

In the field of civil engineering, machine vision has been broadly applied for tracking site elements and measuring their quantities thanks to its cost-effectiveness, and ease-of-use. Examples include, but not limited to, Yang et al. (2010), Park and Brilakis (2012), and Chi and Caldas (2011). These efforts have greatly advanced

vision tracking applications in construction. However, very limited works thus far have been undertaken in human body motion detection, which is the key in applying machine vision to the CTD assessment on construction site. Similar works were undertaken by Li and Lee (2011) and Han et al. (2012). In their works, construction safety analyses were carried out in binocular videos and emphases were placed on fatal accidental monitoring such as fall from ladders. A class of methods also tried Kinetic to detect human joints or body segments (Warade et al. 2012; Wang et al. 2012). The main limitation is that Kinetic operates only in short ranges (< 4 m), leading to this technology incapable for large scale scenes such as construction sites.

PROBLEM STATEMENT AND RESEARCH OBJECTIVE

Current techniques for assessing the CTD risk on construction site involve many manual and subjective procedures. Thus, the results can be error-prone. A more advanced biomechanical analysis tool is desired for reliable assessment of CTD risks. At construction sites, video surveillance already exists and has been routinely used to oversee site activities and record project progresses. The objective of this research is to utilize the video surveillance readily available on site, and investigate a video-based approach that is able to assess the CTD risk among construction workers in an automatic, objective and continuous manner.

AUTOMATED CTD RISK EVALUATION: VIDEO-BASED APPROACH

Figure 1 provides the main workflow of the proposed method. The premise is that the video surveillance is able to "see" the site workers. The technical details of the proposed method are presented in the following steps.

Step 1: Automated body motion detection. The proposed method starts with automatically detecting human body motions of the site workers. A machine vision-based algorithm is proposed to perform this task, which comprises three basic steps: 1) initialization, 2) tracking, and 3) pose estimation. Initialization of vision-based human motion capture usually requires definition of a humanoid model characterizing the shape, appearance, kinematic structure, and initial pose of the object to be tracked (Moeslund et al. 2006). In comparison with other personnel, construction workers can be defined as a person wearing safety gear including safety vests and hard hats (Park and Brilakis 2012). Also, site workers always involve pose changes as a job activity starts. As a result, blob extraction, histograms of oriented gradients (HOG), and color histograms and k-nearest neighbors are applied to detect the site worker's foreground map in video frames. Tracking in visual analysis of human motion is denoted as the process of associating the detected humans in the current frame with those in the previous frames, resulting in a temporal trajectory through the image space (Moeslund et al. 2006). To this end, the kernel-based method is proposed to track the workers on site, as tested in (Park et al. 2011), kernel-based methods had more stable and accurate results than the point-based and contour-based methods. Pose estimation refers to the process of detecting body joint positions or body segments of a human in the resulting regions of the video frames. Currently, two classes of methods are used mostly to estimate the human pose configurations: exemplar based and model based.

Exemplar based methods exploit the "shape context" as a descriptor for finding correspondences between pointsets (Mori and Malik 2006). Input images are matched to each stored view. If there is a stored view sufficiently similar in configuration and pose, the correspondence process succeeds and the locations of the body joints are transferred from the exemplar view to the test shape. The advantage of exemplar based methods is that no key points are necessary. However, exemplar based methods have to search through a huge exemplar database, leading to high complexity for video applications. Model based methods use pictorial structure models to represent a human by a collection of parts (torso, head, upper arm, lower arm, upper leg, and lower leg) in a deformable configuration (Felzenszwalb and Huttenlocher 2005). These parts are separately modeled and interconnected to control the placement of the various parts of the model. By training the model from training examples, humans in an image can be recognized and their constituent parts can be located in the image. In contrast to exemplar-based methods, model-based methods are more efficient.
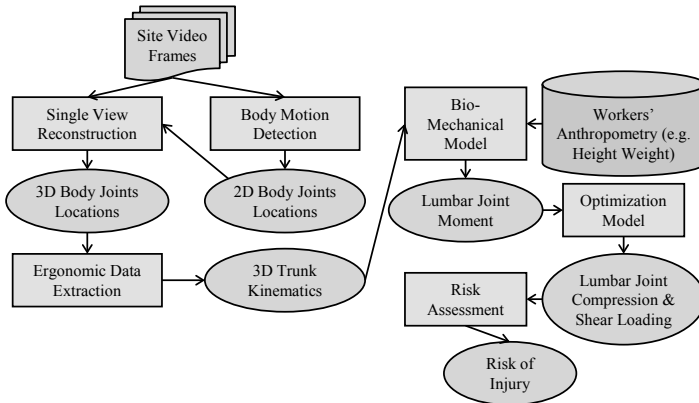
Figure 1: Workflow of proposed methodology

     Step 2: Single view reconstruction. The detected joint locations are in 2D scope. Given that the surveillance camera's location (3D position and 3D orientation) is known, the 3D joint locations are then estimated using single view reconstruction.

     Step 3: Ergonomic data extraction. From Step 2, 3D locations of at least ten (neck, lumbosacral joint, shoulder (bilateral), wrists (bilateral), knee (bilateral) and ankle (bilateral)) joints are obtained. Based on these joint locations, the trunk flexion angle is calculated and the body segment (head, trunk, arm, leg) centers of mass (COM) are estimated in each frame. Next, the linear and angular velocity and acceleration of each body segment COM with respect to the lumbosacral joint are calculated.

     Step 4: Lumbar joint moment determination. In order to estimate spinal (lumbosacral joint) loading, the instantaneous lumbar joint moment is calculated with the ergonomic data obtained from Step 3. Equation (1) provides the procedure of calculating lumbar joint moment, in which a total number of n body segments including trunk, left arm, right arm, left hand load, and right hand load are considered. Variables α_i, β_i and m_i, denote the linear acceleration, angular acceleration and total

mass of the segment i respectively; d_i represents the horizontal distance between the COM of segment i to the lumbar joint and g represents the gravity.

$$Moment_{Lumbar} = \sum_{i=1}^{n} m_i \times (\alpha_i + g) \times d_i + \sum_{i=1}^{n} m_i \times d_i^2 \times \beta_i \quad (1)$$

Step 5: Lumbar joint loading estimation. This step is to estimate lumbar joint loading in which lumbosacral joint compression and shear forces are estimated. The magnitude of spinal compression and shear forces are highly associated with the risk of low back injury. They are determined by the weight of the trunk as well as forces exerted from all trunk tissues (e.g., muscles, ligaments). However, without empirical measurements of lumbar tissue forces, it is not feasible to calculate spinal compression forces due to the redundancy in equations (i.e., there are more unknown tissue forces than equilibrium equations). To obtain spinal compression and shear forces with the use of only trunk anthropometric and kinematics data, optimization models were developed. A well-established double-objective function optimization model (Bean et al. 1988) can be used in estimating lumbar tissue forces and thereby the lumbosacral joint compression and shear forces.

Step 6: Risk assessment. In the current study, both acute and cumulative (long term) LBD risks are considered. Waters et al., (1993) identified the maximum acceptable spinal compressive force of 3400 N. This value is used as a criterion in finding highly risky tasks that could cause acute back injury. To address cumulative risk of injury, an existing risk assessment tool - NIOSH lifting equation (Waters et al. 1993) can be adopted in identifying the risk of awkward trunk postures, repetitive bending motion, and long working hours. As an output of the proposed method, the lifting index calculated using the NIOSH lifting equation will be used to categorize the level of cumulative LBD risks.

IMPLEMENTATION AND PRELIMINARY RESULTS

Considering that prior works exist (Park et al. 2011; Park and Brilakis 2012), this study does not intend to test site worker detection and tracking at the first stage; instead, preliminary experiment has been conducted to validate the performance of 2D pose estimation in the industrial ergonomics lab at WVU (Figure 2). A Canon VIXIA HF M50 camera was used to collect the video data. The camera was fixed on a tripod facing perpendicularly the human subject (Figure 2), configured with its resolution set to be 1.5 MB. As this study focused on LBDs, six groups of typical dynamic bending motions were designed: sagittal symmetric bending (45°), sagittal symmetric bending (90°), asymmetric (45° left) bending (45°), asymmetric (45° left) bending (90°), asymmetric (45° right) bending (45°), and asymmetric (45° right) bending (90°) as shown in Figure 3. These motions all started with the subject standing upright facing the camera and then mimicking the bending motions that take place on construction sites. To the best knowledge of the authors, the designed testing pose configurations have great possibility to cause the occurrence of LBDs.

The estimation of human pose configurations in visual data by exemplar based methods are computational expensive. Thus, this study focused on testing the model based method, which was implemented in Matlab R2012a using the algorithm created by Eichner et al. (2012). At the phase the ongoing research was reported, testing that had been completed included validation of static frames, which were randomly

extracted from the collected video. Each group took 20 frames. Assuming that the subject has been detected, the pose estimation code worked on the region of the subject specified by a bounding box. The result was the body segment detection including head, torso, upper, lower arms and legs as shown in Figure 4. The metric is correct estimation rate. We counted the number of detected body segments that were correctly overlaid on the subject and then calculated the percentage of the correct number over overall detected segments. Table 1 shows the results, from which we concluded the performance was reasonable though the clutter was overlapped with the subject. We observed that the symmetric bending achieved better results than the asymmetric bending. From the results, it was also noted that as the degree of bending increases, the pose detection performance becomes worse. The possible reason would be occlusion of body parts as the subject bended that confused the algorithm. To address this issue, further study is needed, which may utilize the temporal information of the video to track and adjust the body segment trajectory.
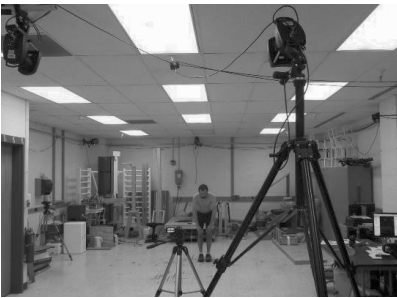
Figure 2: Industrial ergonomics lab

Figure 3: Designed bending postures
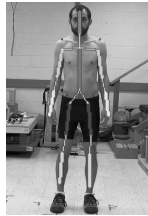
Figure 4: Body segment detection

Table 1. Experimental results

	Sagittal symmetric bending		Asymmetric (45° left) bending		Asymmetric (45° right) bending	
Bending degree	(45°)	(90°)	(45°)	(90°)	(45°)	(90°)
Ave. rate (%)	84	71	75	64	74	62
Ave. time (sec)	4.3	4.0	3.5	3.6	3.7	3.6

CONCLUSION

This paper presented the ongoing research conducted at WVU, which focused on utilizing video surveillance to estimate the risk of CTDs that take place on

construction sites. The main workflow of the methodology was described. This was followed by the preliminary experiment which validated the 2D pose estimation in lab. Results of the study demonstrated the potential of the work. However, there are a number of problems remaining such as how to utilize temporal information of video to increase detection accuracy, real environment testing, joint location extraction and evaluation, and 3D joint recovery. These will be the future work of this study.

ACKNOWLEDGMENTS

REFERENCES

Bean, J.C., Chaffin, D.B., and Schultz, A.B. (1988). "Biomechanical model calculation of muscle contraction forces: A double linear programming method." *Journal of Biomechanics*, 21, 59-66.

Boschman, J.S., van der Molen, H.F., Sluiter, J.K., and Frings-Dresen, M.H. (2012). "Musculoskeletal disorders among construction workers: a one-year follow-up study." *BMC Musculoskeletal Disorders*, 13:196.

Bureau of Labor Statistics. (2011). "Nonfatal occupational injuries and illnesses requiring days away from work, 2010." <www.bls.gov/news.release/archives/osh2_11092011.pdf> (Nov. 10, 2012).

Bureau of Labor Statistics (BLS). (2012). "Nonfatal occupational injuries and illnesses requiring days away from work, 2011." <http://www.bls.gov/news.release/archives/osh2_11082012.pdf> (Nov. 10, 2012).

Chang, C.C., Hsiang, S., Dempsey, P.G., McGorry, R.W., (2003). "A computerized video coding system for biomechanical analysis of lifting tasks." *International Journal of Industrial Ergonomics* 32(4), 239-250.

Chi, S., and Caldas, C.H. (2011). "Automated object identification using optical video cameras on construction sites." *Comp. Aided Civ. Infrastr. Eng.*, 26(5), 368-80.

Coenen, P., Kingma, I., Boot, C.R.L., Faber, G.S., Xu, X., Bongers, P.M., Van Dieen, J.H., (2011). "Estimation of low back moments from video analysis: a validation study." *Journal of Biomechanics*. 44(13): 2369-2375.

Eichner, M. and Marin-Jimenez, M. and Zisserman, A. and Ferrari, V. (2010). "Articulated human pose estimation and search in (almost) unconstrained still images." *ETH Zurich, D-ITET, BIWI, Technical Report No.272*.

Felzenszwalb, P., and Huttenlocher, D. (2005) "Pictorial structures for object recognition." *International Journal of Computer Vision*, 61(1), 55 - 79.

Han, S., Lee, S., and Peña-Mora, F. (2012). "A machine-learning classification approach to automatic detection of workers' actions for behavior-based safety analysis." *ASCE IWCCE*, Clear Beach, FL.

Hignett, S., McAtamney, L., (2000). Rapid entire body assessment. *Applied Ergonomics*, 31(2), 201-205.

Inyang, N., Al-Hussein, M., El-Rich, M., and Al-Jibouri, S. (2012). "Ergonomic analysis and the need for its integration for planning and assessing construction tasks." *J. Constr. Eng. Manage.*, 138(12), 1370–1376.

Kingma, I., De Looze, M.P., Toussaint, H.M., Klijnsma, H.G., Bruijnen, T.B.M., (1996). Validation of a full body 3-D dynamic linked segment model. *Human Movement Science*. 15(6), 833-860.

Kisner, S. and Fosbroke, D. E. (1994). "Injury hazards in the construction industry." *Journal of Occupational Medicine*, 36: 137–143.

Li, C., and Lee, S. (2011). "Computer vision techniques for worker motion analysis to reduce musculoskeletal disorders in construction." *ASCE IWCCE*, Miami, FL.

Marras, W.S., Granata, K.P., (1997). The development of an EMG-assisted model to assess spine loading during whole-body free-dynamic lifting. *Journal of Electromyography and Kinesiology*. 7: 259-268.

McAtamney, L., Corlett, EN., (1993). RULA – A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2), 91-99.

Moeslund, T. B., Hilton, A., and Kru, V. (2006). "A survey of advances in vision based human motion capture and analysis." *Computer Vision and Image Understanding*, 104(2), 90–126.

Mori, G., and Malik, J. (2006). "Recovering 3d human body configurations using shape contexts." *IEEE Trans. PAMI*, 28(7), 1052-1062.

National Institute of Occupational Safety and Health (NIOSH). (2007). *Simple Solutions Ergonomics for Construction Workers*, <www.cdc.gov/niosh/docs/2007-122/pdfs/2007-122.pdf> (Nov. 14, 2012).

North Carolina Department of Labor (NCDL). (2012). Sample program on "Construction Safety and Health Policy". <www.nclabor.com/osha/consult/sampleprograms/ConstructionS&HPolicy.pdf> (Nov. 14, 2012).

Occupational Safety and Health Administration (OSHA). (2012). *White Paper on Injury and Illness Prevention Programs*, <http://www.osha.gov/dsg/topics/safetyhealth/OSHAwhite-paper-january2012sm.pdf> (Nov. 15, 2012).

Ohio Bureau of Workers' Compensation (OHBWC). (2012). Ergonomics Best Practices for the Construction Industry, <www.ohiobwc.com/downloads/brochureware/publications/ConstSafeGrant.pdf> (Nov. 14, 2012).

Park, M.W., Makhmalbaf, A., and Brilakis, I. (2011). "Comparative study of vision tracking methods for tracking of construction site resources." *Autom. Constr.*, 20(7), 905-915.

Park, M.W., and Brilakis, I. (2012). "Construction worker detection in video frames for initializing vision trackers." *Autom. Constr.*, 28, 15–25.

Wang, J., Liu, Z., Chorowski, J., Chen, Z., and Wu, Y. (2012). "Robust 3D action recognition with random occupancy patterns." *ECCV 2012, LNCS*, 872-885.

Warade, S., Aghav, J., Claude, P., and Udayagiri, S. (2012). "Real-time detection and tracking with Kinect." *Intl. Conf. Comp. Info. Tech.*, Bangkok, 86-89.

Waters, T.R., Putz-Anderson, V., Garg, A., and Fine, L.J., (1993). "Revised NIOSH equation for the design and evaluation of manual lifting tasks." Ergonomics, 36(7), 749-776.

Yang, J., Arif, O., Vela, P. A., Teizer, J., Shi, Z. (2010). "Tracking multiple workers on construction sites using video cameras." *Adv. Eng. Inf.*, 24(4), 428-434.

Applying Regression Analysis to Predict and Classify Construction Cycle Time

M. F. Siu[1], R. Ekyalimpa[2], M. Lu[3] and S. Abourizk[4]

[1]Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada; PH (780) 655-8532; email: siumingfungfrancis@gmail.com
[2]Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada; PH (780) 492-3496; email: rekyalimpa@ualberta.ca
[3]Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada; PH (780) 492-5110; email: mlu6@ualberta.ca
[4]Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada; PH (780) 492-8096; email: abourizk@ualberta.ca

ABSTRACT

Regression techniques are commonly used for addressing complicated prediction and classification problems in civil engineering thanks to its simplicity. For a given dataset, the linear regression from the input space to the output variables can be achieved by using the "least square error" approach, which minimizes the difference between the predicted and actual outputs. The "least mean square" rule can also be used as a generic approach to deriving solutions on linear or non-linear regressions. The paper addresses the fundamental algorithms of "least square error" and "least mean square" in order to facilitate the prediction and classification of cycle times of construction operations. The classic XOR problem is selected to verify and validate their performances. A viaduct bridge was installed by launching precast girders with a mobile gantry sitting on two piers. The effectiveness of regression techniques in classifying and forecasting the cycle time of installing one span of viaduct considering the most relevant input factors in connection with operations, logistics and resources are demonstrated.

INTRODUCTION

Classification and predication play an important role to the success of planning and control of a construction project. The production rate tracking can be beneficial to forecasting project performances such as the expected activity and project completion times. Corrective actions or changes can be made to tackle any adverse impact on schedule. Yet, applications of artificial intelligence are hampered by its complexities. The operations simulation and artificial neural computing are the two main streams of research which have focused on cycle-time and production rate predictions of construction operation processes. Previous research efforts by Teicholz

(1963), Ahuja and Nandakumar (1985), Halpin (1977), Lu (2003), AbouRizk (2010) have greatly contributed to advancing production rate predictability by implementing simulation techniques. Contributions are claimed to have better planning and control during the project executions, such as estimating the project duration and cost under uncertainty, productivity improvement and potential savings in time or cost (Halpin, 1977, Pritsker et al. 1989, Sawhney and AbouRizk 1995, AbouRizk and Mohamed 2000, Lu 2003 and Tian et al. 2010).

The regression modeling, formally named as linear regression analysis, is profoundly addressed in this paper, which is the earliest prototype that heralded the later development of artificial neural networks. Linear regression techniques are widely applied in the construction field, such as Thomas and Sakarcan (1994); Sanders and Thomas (1993). The basic mechanism is to determine the output y by multiplying each input variable x and an associated weight w as shown in Eq. (1). The weights can be determined by using historical records. As linear regression evolved into neural network computing, research efforts such as Mutlu al et. (2008) used artificial neural networks to forecast the daily water flow at multiple gauge stations in the agricultural domain; Portas and AbouRizk (1997), Lu et al. (2000), embedded fuzzy logic into neural computing and implemented sophisticated probabilistic inference neural networks to estimate labor production rates, respectively. The neural network computing research has mainly focused on advancing methodologies (single- and multi-layer perceptrons, radial-basis functions and support vector machines) in terms of computing efficiency and effectiveness (Haykin, 1998).

However, sufficient knowledge of neural network computing is required on construction professionals in order for them to fully trust and harness these advanced tools instead of simply using them as "black box". This research has contributed to elucidating on fundamental regression analysis techniques underlying neural network computing, namely: (1) the "least square error" and (2) the "least mean square" algorithms, in an effort to facilitate prediction and classification applications in construction engineering. The following sections provide mathematical background, algorithmic verification using the XOR dataset, and implementation of regression techniques to predict and classify precast bridge segment erection cycle times.

$$y = w_0 + w_1 x_1 + w_2 x_2 + \cdots \tag{1}$$

LEAST SQUARE ERROR APPROACH

The least square error approach is the earlier form used for numerical regression and prediction. The general equation of linear regression model is in the form of Eq. (1). The weight parameters (w_n) could be optimized and stabilized by analyzing available input data (x_n). The analytical matrix-approach transforms Eq. (1) in forms of vector as given in Eqs. (2) and (3). The error is defined as the difference between the actual output y and the model's output (summation of xw.) By least square adjustment techniques, the error of the system can be minimized by taking its partial derivatives with respect to w_n which are set as zero to derive optimal solutions, as shown in Eqs. (4) to (6). The Eqs. (7) to (9) show a system of rearranged equations which can be expressed in a compact matrix form as Eq. (10). The weight w parameters can be easily calculated by matrix manipulations only involving

parameters of input x and output y. It is noteworthy that w_0 in Eq. (1) is the error (noise or disturbance) term commonly defined in applied statistics. This term is essential to account for any unobserved random variable that presents noise to the linear relationship between input x and output y. According to Harrell (2001), the error term should be statistically independent and identically distributed, approximately normally distributed and have a common variance. To prove the effect of unobserved random variables on the linear regression model, Eq. (1) can be interpreted as Eq. (11). The linear regression system explicitly considers the inputs from 1 to p while factoring in both observed inputs (1 to p) and unobserved inputs (p to ∞). The indispensable bias term maintains the integrity of regression techniques resulting in a unique solution. In contrast, the "least mean square rule" approach utilizes iterative procedures to stabilize the weights for each input variable along with the bias, eventually minimizing errors on model outputs.

$$W = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}; \ X = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} \tag{2}$$

$$y = f(X,W) = X^T W \tag{3}$$

$$\frac{\partial \sum \text{error}^2}{\partial w_0} = \sum 2(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2}))(-1) = 0 \tag{4}$$

$$\frac{\partial \sum \text{error}^2}{\partial w_1} = \sum 2(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2}))(-x_{i1}) = 0 \tag{5}$$

$$\frac{\partial \sum \text{error}^2}{\partial w_2} = \sum 2(y_i - (w_0 + w_1 x_{i1} + w_2 x_{i2}))(-x_{i2}) = 0 \tag{6}$$

$$\sum y_i = n w_0 + w_1 \sum x_{i1} + w_2 \sum x_{i2} \tag{7}$$

$$\sum x_{i1} y_i = w_0 \sum x_{i1} + w_1 \sum x_{i1} x_{i1} + w_2 \sum x_{i2} x_{i1} \tag{8}$$

$$\sum x_{i2} y_i = w_0 \sum x_{i2} + w_1 \sum x_{i2} x_{i1} + w_2 \sum x_{i2} x_{i2} \tag{9}$$

$$W = (X^T X)^{-1} X^T Y \tag{10}$$

$$y_i = \sum_{j=1}^{\infty} w_j x_{ij} = \sum_{j=1}^{p} w_j x_{ij} + \sum_{j=p}^{\infty} w_j x_{ij} = \sum_{j=1}^{p} w_j x_{ij} + w_0 \tag{11}$$

LEAST MEAN SQUARE APPROACH

The least mean square rule expresses the partial derivative in terms of vector $\bar{G}$ defined as the system error in linear regression analysis, as given in Eq. (12). The least mean square rule represents a generic approach to optimize the weights by continuously seeking stabilized value of the bias w_0 expressed as Eq. (13). The objective is to progressively evaluate w by applying iterative procedures aimed at minimizing the system error, as given in Eqs. (14) to (17). $Y(i)$ in Eq. (14) denotes the output. The error in relation to each input pattern can be computed by subtracting the desired output $d(i)$ and $Y(i)$ as Eq. (15). W in Eq. (16) and w_0 in Eq. (17) can be updated by using the learning rate parameter η until the values are steady, when the optimum solution of $W(i)$ is reached. The range of η can be estimated by the

eigenvalue of the correlation matrix X. Generally, more time is needed to search the optimum solution if the step-size chosen is too small, while no solution can be reached if the value is too large due to increased chances of divergence. As the data may be highly linearly-non-separable, pre-processing data is essential to transform the data from being linearly-non-separable to being linearly-separable.

$$\bar{G} = \frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_o} \frac{\partial y_o}{\partial w_{ij}} = -(d_o - y_o)y_i \tag{12}$$

$$W(\text{new}) = W(\text{old}) - \eta \frac{\partial E}{\partial w_{ij}} \tag{13}$$

$$Y(i) = W(i) \times X(i) + w_0(i) \tag{14}$$

$$Err(i) = d(i) - Y(i) \tag{15}$$

$$W(i+1) = W(i) + \eta \times Err(i) \times X(i) \tag{16}$$

$$w_0(i+1) = w_0(i) + \eta \times Err(i) \tag{17}$$

DATA PRE-PROCESSING

Data can be difficult to be classified if the input data of the problem are characteristic of high non-linearity. The XOR (eXclusive OR) problem is used to illustrate the "non-linearity" problem. The inputs and outputs are shown in Table 1. Figure 1 graphically plots the four corresponding points; no straight line can be drawn to cluster the points having the same outputs. Thus, direct classification based on linear regression is not feasible owing to the non-linearity in data. To tackle the problem, the raw data is firstly normalized. K-means clustering technique is employed to evaluate the distances between the input pattern and the center of a cluster r_i. Note that the K-means clustering algorithm is a special case of self-organizing maps (Haykin, 1998). The P-nearest neighbor algorithm is then applied to determine the sigma σ of the Gaussian function for each cluster. Finally, the data is transformed by using the Gaussian function as shown in Eq. (18).

The K-means clustering and P-nearest neighbor algorithms are implemented considering two clusters on the XOR dataset. The two clustering centers coordinates are determined as (0, 0), (1, 1) and σ as 1.414. The transformed inputs are thus obtained by data pre-processing, ready for regression analysis. A straight line can separate the four points with into respective classifications. Figure 2 shows the linearly-separable transformed inputs. Both least square error and least mean square rule approaches were applied, achieving the same results. w_1 and w_2 are evaluated both as 20.438 and w_0 is −31.834. A multivariate regression equation can be expressed in Eq. (19) to represent the complete prediction model, where x_1 and x_2 are the two input parameters.

$$x_i = e^{\frac{-r_j^2}{2\sigma^2}} \tag{18}$$

$$Y_{output} = 20.438\, e^{\frac{-[(X_1-0)^2+(X_2-0)^2]}{-2(1.414)^2}} + 20.438\, e^{\frac{-[(X_1-1)^2+(X_2-1)^2]}{-2(1.414)^2}} - 31.834 \tag{19}$$

Table 1. XOR Inputs and Outputs

Inputs Before Transformation			Inputs After Transformation		
x_1	x_2	y	x_1	x_2	y
0	0	1	1.000	0.607	1
0	1	0	0.779	0.779	0
1	0	0	0.779	0.779	0
1	1	1	0.607	1.000	1
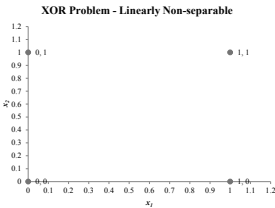
Figure 1. Linear Non-separable
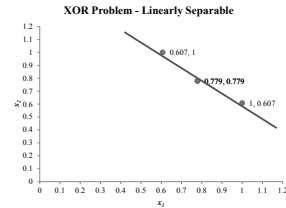
Figure 2. Linear Separable

BRIDGE SPAN ERECTION CASE STUDY

A bridge construction project is used to demonstrate the effectiveness in implementing cycle-time prediction and classification by using regression analysis techniques. The bridge is a new artery linking Hong Kong and Shenzhen, China, and consists of 227 post-tensioned spans of viaduct. A typical span is made up of 14 precast segmental box girders (12m×2.5m×2.8m of each). The stepping girder precast installation method was used to accelerate the viaduct construction process (Chan and Lu, 2009). The precast segments were fabricated near Shenzhen and hauled to the site for installation. The site was too congested to keep all segments in the convenient proximity of the site crew. As an alternative, the precast segments were partially stored in a remote storage area and transported to the working span by trailer trucks, without any intermediate storage or buffer.

In order to assist the contractor in deciding on how many precast segments to be placed at the remote storage area and how far away to locate the remote storage area while maintaining the target one-span erection cycle time of four and half working days, four input factors relevant to site operations and logistics planning were identified and assessed, namely: (1) the number of trailer trucks rented for hauling segments (the site only considered the options of two trailer trucks or three), (2) one-batch or two-batch precast segment delivery modes (14 segments can be delivered either in one batch on the night before installation operations starts or in two batches, which means the first batch of seven segments would be delivered on the night before installation starts and the second batch delivered on the following night), (3) the percentage of the total number of segments on one span to be placed in the remote storage area and (4) the haul duration for a trailer truck to transit from the remote storage area to the working span. Table 2 shows the 30 cycle-time records resulting from simulations, which define 30 different scenarios.

Table 2. Viaduct Installation Cycle-Time Records for Regression Analysis

Rec ID	No. of Tracker	Delivery Batch	Segment at RSA (%)	Duration to RSA	Desired Install Hours	*Predicted Install Hour*	Desired Prod Class	*Predicted Prod Class*
1	2	1	0.00	0.00	103.61	-	0	0
2	3	1	0.50	0.50	104.76	-	0	0
3	3	1	0.29	0.33	104.76	-	0	0
4	3	1	0.50	0.33	104.78	-	0	0
5	3	1	1.00	0.50	105.78	-	0	0
6	3	1	1.00	0.33	105.78	-	0	0
7	3	1	0.29	0.75	108.38	-	1	**0**
8	3	1	0.50	0.75	109.36	-	1	1
9	2	1	0.50	0.50	111.51	-	1	1
10	3	1	1.00	0.75	112.05	-	1	1
11	3	1	0.71	0.75	112.41	-	1	1
12	2	1	0.29	0.75	112.72	-	1	1
13	2	1	1.00	0.50	114.15	-	1	1
14	2	1	0.50	0.75	115.70	-	1	1
15	2	1	0.71	0.75	116.47	-	2	2
16	2	1	1.00	0.75	116.61	-	2	2
17	2	2	0.29	0.50	116.67	-	2	2
18	2	2	1.00	0.33	116.70	-	2	2
19	2	2	1.00	0.75	116.71	-	2	2
20	2	2	0.00	0.00	116.74	-	2	2
21	2	2	0.57	0.50	116.74	-	2	2
22	2	2	0.57	0.75	116.74	-	2	2
23	3	2	0.57	0.50	116.74	-	2	2
24	3	2	0.57	0.75	116.74	-	2	2
25	3	1	0.29	0.50	104.76	-	0	0
26	3	1	0.71	0.50	104.89	-	0	0
27	3	1	0.71	0.33	104.89	-	0	0
28	3	1	0.00	0.00	105.77	-	0	0
29	2	1	0.71	0.33	106.00	-	0	0
30	2	1	0.29	0.50	108.47	-	1	**0**
31	2	2	0.29	0.75	N/A	118.118	2	2
32	3	2	1.00	0.50	N/A	114.425	2	2
33	2	2	1.00	0.50	N/A	118.010	2	2
34	3	2	1.00	0.75	N/A	114.763	2	2
35	2	1	0.29	0.33	N/A	104.030	0	0
36	2	1	0.50	0.33	N/A	105.373	0	0
37	2	1	1.00	0.33	N/A	110.371	1	1
38	2	1	0.71	0.50	N/A	111.135	1	1
39	2	2	0.29	0.33	N/A	115.589	2	2
40	2	2	0.57	0.33	N/A	116.276	2	2

To facilitate the implementation of the algorithms, data normalization, K-means clustering analysis and P-nearest neighbor algorithm were programmed to normalize the non-linear data. Both least square error and least mean square algorithms were coded. It is noteworthy that the number of cluster centers was estimated before executing the K-means clustering algorithm. Cluster center initialization is controlled by a random seed in the computer program. One proposed solution is to evaluate the root mean square error (RMSE) by using n desired and computed outputs given one particular random seed, as shown in Eq. (20). The random seed leading to the smallest RMSE will then be used in the regression analysis. The accuracy of predictions significantly depends on the selection of number of cluster centers, and the quantity and quality of available datasets. In general, the larger size of the dataset, the higher quality of the dataset, the higher

accuracy the predictions. The smallest root mean square error generated is 1.582 hour (Figure 3) by trying 100 random seeds on 30 records ($n = 30$). Ten "unseen" scenarios were predicted by using the derived multivariate regression equation ("predicted install hour" column with ID 31 to 40 in Table 2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_{computed,i} - x_{desired,i})^2}{n}} \qquad (20)$$
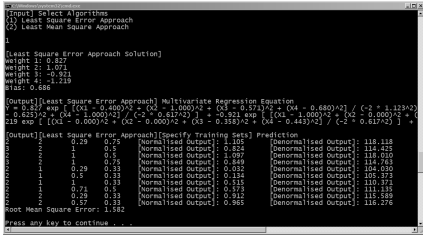
Figure 3. Cycle Time Prediction
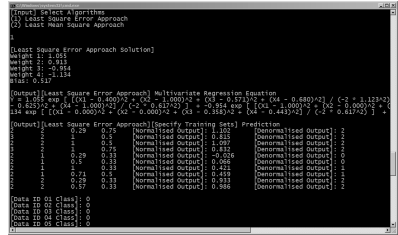
Figure 4. Productivity Classification

Productivity classification is beneficial to decision making during project execution. The contractor defined three classes based on span installation hours (less than 108 hours is "high", between 108 hours and 116 hours is "medium" and longer than 116 hours is "low"). The 30 data set (ID 1 to 30) were chosen to determine the weights of the regression equation. Figure 4 shows the classification results with RMSE 0.327. Only two out of thirty records (ID 7 and 30) are incorrectly classified ("predicted prod class" differing from "desired prod class"). Input patterns can be generalized based on classification results: In order to achieve high productivity, segments must be delivered in one batch, less than 29% segments are stored at remote storage area, and the trailer truck transit time must be within half an hour.

Table 3: Data Manipulations of the Predicted Productivity Classes

Prod Class	No. of Tracker	Delivery Batch	Segment at RSA (%)	Duration to RSA
0 (Class 1 – High)	2 to 3	1	0.00 – 1.00	0.00 – 0.75
1 (Class 2 – Med)	2 to 3	1	0.29 – 1.00	0.50 – 0.75
2 (Class 3 – Low)	2 to 3	1 to 2	0.00 – 1.00	0.00 – 0.75

CONCLUSION

Regression techniques are effective to analyze construction operations in terms of cycle-time prediction and productivity classification. The fundamentals of the "least square error" and the "least mean square" algorithms are contrasted and clarified. Both techniques are verified and validated by using the classic XOR problem and a bridge construction project. Also, the "best fit" classification model is obtained through assigning the number of cluster centers by trial and error. Analytical methods or guidance on how to set up the clusters given a particular dataset can be further generalized.

REFERENCES

AbouRizk, S. 2010. "The role of simulation in construction engineering and management." *Journal of Construction Engineering and Management*, ASCE, 136 (10), 1140–1153.

AbouRizk, S. and Mohamed, Y. 2000. "Simphony-an integrated environment for construction simulation." *Proceedings of Winter Simulation Conference, Orlando*, FL , USA. 1907–1914.

Ahuja, N. T. H., and Nandakumar, V. 1985. "Simulation model to forecast project completion time." *Journal of Construction Engineering and Management*, ASCE, 111(4), 325–342.

Chan, W. H. and Lu, M. 2009 *Artificial intelligence-integrated construction simulation method*, VDM Verlag, Saarbrücken, Germany.

Halpin, D. W. 1977. "CYCLONE - A Method for Modeling Job Site Processes." *J. Constr. Div.*, ASCE. 103(3), 489–499.

Harrell F. E. 2001. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*, Springer.

Haykin, S. 1998. *Neural Networks: A Comprehensive Foundation*, Prentice Hall.

Thomas, R. and Sakarcan, A. 1994. "Forecasting Labor Productivity Using Factor Model." *Journal of Construction Engineering and Management*, 120(1), 228–239.

Teicholz, P. 1963. *A Simulation Approach to the selection of Construction Equipment*. Technical Report No. 26, The Construction Institute, Stanford University.

Tian, X., Mohamed, Y. and AbouRizk, S. 2010. "Simulation-based aggregate planning of batch plant operations". *Canadian Journal of Civil Engineering*, 37(10), 1277–1288.

Lu, M. (2003), "Simplified Discrete-Event Simulation Approach for Construction Simulation", *Journal of Construction Engineering and Management*, ASCE, 129(5), 537–546.

Lu, M., AbouRizk, S. and Herman, U. 2000. "Estimating Construction Productivity using Probability Inference Neural Network". *Journal of Computing in Civil Engineering*, 14(4), 241–248.

Portas, J. and AbouRizk, S. 1997. "A Neural Network Model for Estimating Construction Productivity." *Journal of Construction Engineering and Management*, ASCE, 123(4), 399–410.

Pritsker, A., Sigal, C. and Hammesfahr, R. 1989. *SLAM II Network Models for Decision Support*. Prentice-Hall, Englewood Cliffs, N. J.

Mutlu, E, Chaubey, I., Hexmoor, H. and Bajwa, S. G. 2008. "Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed", *Hydrological Processes*, 22, 5097.

Sanders, S. R. and Thomas, H. R. 1993. "Masonry productivity forecasting model", *Journal of Construction Engineering and Management*, ASCE, 119(1), 163–179.

Sawhney, A. and AbouRizk, S. 1995. "HSM–Simulation-based Project Planning Method for Construction Projects." *Journal of Construction Engineering and Management*, ASCE, 121(3), 297–303.

Mobile LiDAR Data Collection and Analysis for Post-Sandy Disaster Recovery

Jie Gong

Assistant Professor, Dept. of Civil & Environmental Engineering, Rutgers, the State University of New Jersey, Phone: 848-445-2881, jiegong.cee@rutgers.edu

ABSTRACT

On October 29th, Hurricane Sandy wreaked havoc in the much part of East Coast. The crumbling infrastructure along the East Coast suffered major damage during this historical storm, leaving tens of millions of residents without electricity. Rapidly acquired high-quality and high-resolution spatial and photo data on damaged structures are of great value to scientific studies and post disaster recovery-related decision makings. This paper describes a mobile LiDAR study conducted in the sandy impacted areas. The application and utility of mobile LiDAR data for disaster response and recovery is presented and discussed. The preliminary results of this research demonstrated mobile LiDAR data fill in the gap between airborne remote sensing data and typical ground survey data.

INTRODUCTION

On October 29, 2012, Hurricane Sandy landed on New York and New Jersey, causing nearly $60 billion in damage to the region's property, infrastructure, and commerce. In many parts of New York City and New Jersey, infrastructures have been crippled; beaches and boardwalks have been erased; family homes have been condemned. In the Northeast, hurricanes are not typical weather events – coupled with the region's dense population and intricate transportation network, Sandy's lasting devastation is considered second only to the damage caused by Hurricane Katrina.

Airborne remote sensing technologies, such as aerial imagery and airborne LiDAR, are primary tools for collecting time critical post-disaster geospatial information. For example, on October 30, 2012, the day after Sandy's landing on the east coast, the National Oceanic and Atmospheric Administration (NOAA) conducted an aerial photography survey in the hardest hit areas. The collected aerial imagery showed wide-spread destruction around the costal line, and has become a crucial data source used by federal, state, and local officials to plan search and rescue operations, route personnel and machinery, coordinate recovery efforts, and provide a cost-effective way to better understand the damage sustained to both property and the environment. However, airborne remote sensing has its limitations. The accuracy and resolution of airborne data is typically lower than many ground-based remote sensing technologies. Many building structure evaluation and environmental assessment tasks require high accuracy geospatial data. Indeed, the disaster response and recovery process inevitably involves a large amount of foot-on-ground surveying activities.

These activities are tedious and unsafe in many disaster impacted areas. Most of these activities only provide a local and discrete measurement of building and infrastructure features and/or location and are unable to portray a holistic picture of disaster areas. Once the bulldozer starts rolling, critical information and valuable scientific evidence such as how different types of buildings and infrastructures react to severe weather conditions is lost.

The purpose of this research is to explore the potential of mobile LiDAR, a land-based remote sensing method, to support disaster response and recovery process. Mobile LiDAR is designed for collecting survey grade LiDAR data over large areas that would be impractical with static LiDAR sensors, but require an accuracy and resolution that exceed airborne technologies. Hurricane Sandy caused damage so widespread that it overwhelmed the capacity of traditional surveying methods. The use of mobile LiDAR in such a post-disaster scenario is a very logical choice. Due to the cost, the use of mobile LiDAR for disaster response and recovery is uncommon and related research literature is scarce. This paper describes the planning and execution of a four-day mobile LiDAR study in the Sandy-impacted areas including New York City and New Jersey Shoreline. Use of these collected LiDAR data to support post-disaster recovery is highlighted and discussed. The lessons learned in this data collection effort will provide the foundation for future improvement of this growing technology to better support post-disaster recovery process.

BACKGROUND

Mobile LiDAR (Laser Detection and Ranging) is a growing remote sensing technology that has been used to collect geospatial data along highway systems (Yen et al. 2011; de la Garza et al 2011; Gong et al. 2012; Jalayer et al. 2012). A mobile LiDAR system often consists of cameras and multi-return LiDAR sensors mounted on a vehicular platform whose positions and headings are precisely tracked by fusing data from the onboard GPS and an Inertia Measurement Unit (IMU). The LiDAR sensors typically incorporate lasers that operate at pulse rates up to 1 million Hz, producing dense and highly accurate point clouds along 200-400m-wide swaths; collected at forward speeds of approximately 20-50 MPH. In combination with the laser scanning, imagery is acquired by several digital cameras. The digital camera system is calibrated to the center of the scanner. The system can be mounted on a variety of platforms including road, rail, and marine vehicles. Compared to other remote sensing data such as airborne LiDAR, mobile LiDAR data provides a highly accurate, detailed representation of terrain.

The system used in this study is an Optech LYNX Mobile Mapper M1 system. The LYNX system relies on two 500 kHz LiDAR sensors to collect a million points per second while maintaining survey grade quality precision. The LYNX M1 field of view is 360 degrees (without obstructions) which includes a Real-Time Point Cloud (RTPC) interface, an on-the-fly tool to see the swath of data being collected for field verification. Without ground control points but with a mobile base station, the Optech system is capable of obtaining LiDAR data sufficient for feature extraction of planimetric and topographic features typically at an absolute accuracy of +/-10cm @ 1 σ in good GPS coverage areas and an relative accuracies of +/- 5cm @ 1 σ

anywhere within the project area. The collected LiDAR data obtained will consist of but not limited to X coordinate, Y coordinate, Z coordinate, GPS time, return number, and intensity value.

FIELD DATA COLLECTION

The post-Sandy mobile LiDAR data collection was conducted between December 5 and December 9, 2012. The areas of interest for this study are distributed along Staten Island and Rockaway of New York and the Jersey Shorelines. The areas of particular interest to us along the New Jersey shorelines are: (1) Seaside Heights; (2) Normandy Beach; (3) Ortley Beach; and (4) Mantoloking; all of them are among hardest hit areas in this storm. Figures 1 through 5 show the specific locations of data collection.

(a)   (b)

(c)   (d)
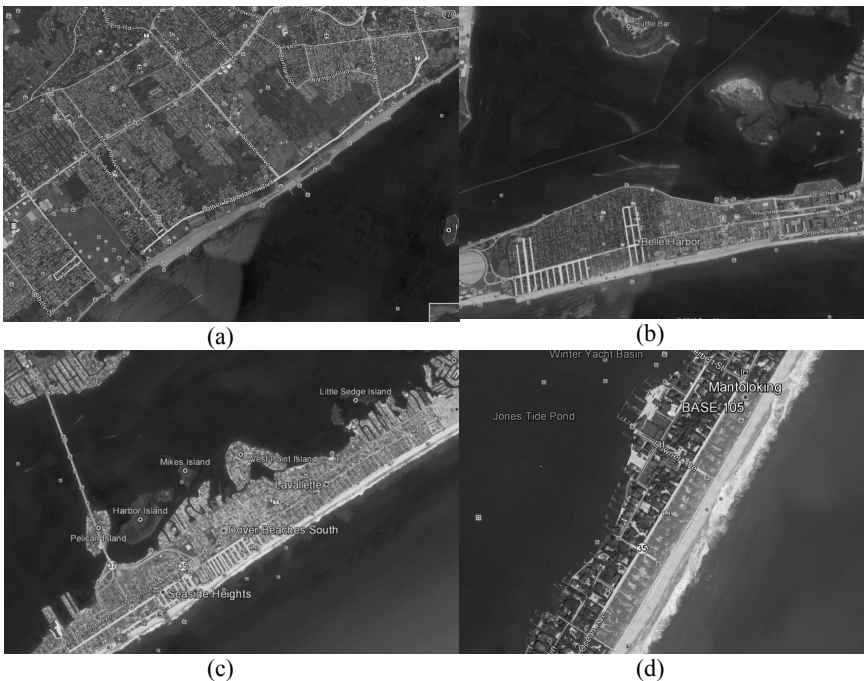
Figure 1. Data Collection Locations in (a) Staten Island; (b) Rockaway; (c) Seaside Heights, Ortley Beach, and Normady Beaches; and (d) Mantoloking

Figure 2. The Mobile LiDAR Data Collection Van

Due to the importance of imagery to this study, the data collection was mostly conducted during the daytime from 9:00AM – 4:00PM. While it is certainly possible to collect data starting at 7:00AM, the sunlight during the time period of 7:00AM to 9:00AM often saturates the cameras' lenses system due to the sunlight angle. Also, in the winter, the daylight time is very short – after 4:00PM, it is too dark to collect any good quality imagery. The busy traffic in the city like New York also has considerable impact on data collection. The combination of these factors created significant constraints to efficient mobile LiDAR data collection. According to our time logging record, the daily data collection rate is approximately 20 miles per hour, which is lower than a normal rate of 30 miles per hour in densely populated urban areas.

DATA PROCESSING AND APPLICATIONS

Once collected, the mobile LiDAR data will be processed in a series steps (Figure 3). These steps are fairly standard post-processing steps used on both airborne and mobile LiDAR data. The detailed explanation of these steps can be found in many existing studies (Graham 2010; Vincent 2010). The focus of this paper is on what the processed data can do for the disaster recovery process. The following details the application of mobile LiDAR data to support post-disaster operations.
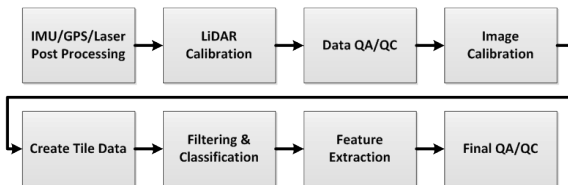
Figure 3. Mobile LiDAR Data Post-processing Steps

Visualization

The high-resolution, three-dimensional images generated by LiDAR can be used to create a virtual reality of the flooded community. The images can be used to

measure the elevation change, the shifting of foundations and houses, the extent of soil erosion, the volume of debris, and so on (Figure 4). Geo-referenced imagery provides another way to visualize the damage and compare it to the pre-disaster condition (available from most map services, such as Google Map and Bing Map) (Figure 5). The three-dimensional images of the flooded community can also be examined in close details by coastal engineers to study wave, surge, and wind damage mechanisms.   Wave and surge models have been shown to result in increasing errors as they move overland arising from two major factors: (A) Increased dissipation overland is not fully accounted for in standard models, particularly for water waves; and (B) Overland wind stress will be partially absorbed by canopy elements like trees and houses and will not entirely reach the water surface, with strong implications for surge and waves (Kennedy et al., 2011). The mobile LiDAR data provides a data set with unprecedented detail and accuracy to support the study of the impact of canopy factors on wave dissipation process. These will provide the increased predictive accuracy and detail during surge events needed to make decisions for sustainability.

(a)

(b)

Figure 4. (a) Sandy Damages at Ortley Beach Colored by Elevation; (b) Sandy-related Soil Erosion at Staten Island

Figure 5. Geo-referenced Sandy Imagery

Flood Scenario Reconstruction and Analysis

By examining debris trace and water markers on structures and trees, the flood scenario can be reconstructed to study the extent of flooding and its impact on city assets. Figure 6 shows the reconstructed flood scenario for a coastal city. Such flood reconstruction provides a system view of the extent of flooding, assisting the identification of vulnerable assets.

Figure 6. Flood Scenario Reconstruction

Engineering Surveying Drawing

The 3D images also can be used to generate survey quality engineering drawings that can serve as the basis for clean-up, demolition, recycling, rebuilding and recovery efforts (Figure 7). It would take several months for traditional surveyors

to survey the disaster areas where the mobile LiDAR has scanned. These survey drawings were developed for the Father Capodanno Blvd. which has sustained considerable damage during the hurricane Sandy. NYC Department of Design and Construction has been pushing hard to develop these post-disaster as-built surveying record. The data collected in this study can be used to quickly generate these drawings for providing baselines for rapid road repair work. This reduces the turnaround time for acquiring survey quality drawings from several months to several days.
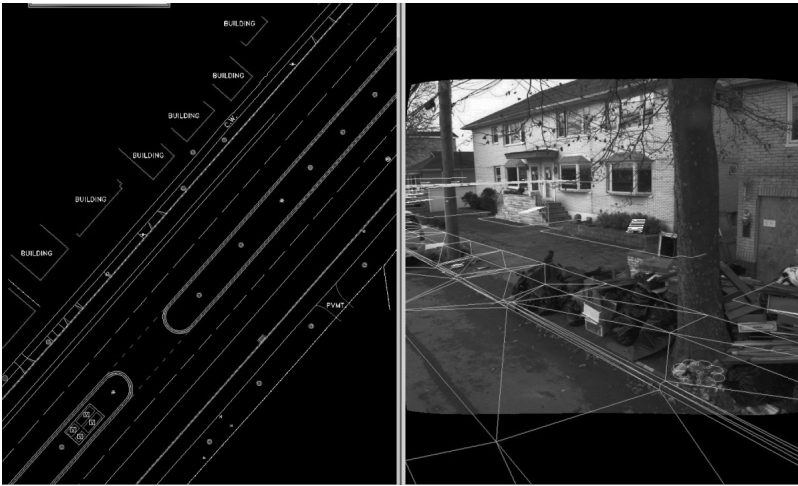
Figure 7. Survey Quality As-Built Drawing Development from Sandy LiDAR and Imagery Data

CONCLUSION

This research explored the utility of Mobile LiDAR for supporting post-Sandy recovery operations. Mobile LiDAR collected data at a resolution and an accuracy that cannot be achieved by airborne remote sensing technologies. While the analysis of the collected LiDAR data is still a largely ongoing processing, the preliminary study shows with mobile LiDAR, disaster assessment can become a quicker process that concurrently uses a survey grade mapping system while producing 3-D condition data. At the time of this writing, multiple state agencies have begun to leverage this dataset for visualizing the extent of damage and planning recovery strategies. It is because the high resolution and high accuracy geospatial data collected by a mobile LiDAR system provide a rich data set that cannot be collected by other technologies. The data set provides new opportunities for wind, wave, and storm surge studies. Several wave and surge studies with the collected data are under development. At last, but not the least, visualization and quantification of disaster impact can be greatly enhanced by the use of mobile LiDAR data. One of the challenges but opportunities for using mobile LiDAR data to support disaster recovery process is that there

appears to be very little literature on best practices and how to tailor LiDAR data processing algorithms for disaster recovery purposes.

REFERENCE

De la Garza, J., Figueroa, C. F., Howerton, C. G., Plummer, J., Roca, I., Shogli, O., Sideris, D., and Uslu, B. (2009). "Implementation of IP-S2 Mobile Mapping Technology for Highway Asset Visualization in Data Collection to Benefit VT-VDOT TAMS Project." Center for Highway Asset Management Programs, Virginia Department of Transportation.

Gong, J., Zhou, H., Gordon, C., and Jalayer, M. "Mobile Terrestrial Laser Scanning for Highway Inventory Data Collection: A Review." 2012 International Workshop of Computing in Civil Engineering, Clearwater, FL.

Graham, L. (2010). "Mobile Mapping Systems Overview." Photogrammetric Engineering and Remote Sensing, March.

Jalayer, M., Gong, J., Zhou, H., and Grinter, M. (2013). "Evaluation of Remote Sensing Technologies for Collecting Roadside Feature Data to Support Highway Safety Manual Implementation." 92nd Annual Meeting of the Transportation Research Board, Washington D.C.

Kennedy, A.B., Rogers, S., Sallenger, A., Gravois, U., Zachry, B., Dosa, M., and Zarama, F. (2011). "Building Destruction from Waves and Surge on the Bolivar Peninsula during Hurricane Ike," J. Waterway, Port, Coastal and Ocean Eng.-ASCE, 137(3), 132-141, doi:10.1061/(ASCE)WW.1943-5460.0000061.

Vincent, R. A., and Ecker, M. (2010). "Light Detection and Ranging (LiDAR) Technology Evaluation." Missouri Department of Transportation.

Yen, K. S., Ravani, B., and Lasky, T. A. (2011). "LiDAR for Data Efficiency." Washington State Department of Transportation.

Effect of Color Space, Color Channels, and Sub-Image Block Size on the Performance of Wavelet-Based Texture Analysis Algorithms: An Application to Corrosion Detection on Steel Structures

M. R. Jahanshahi[1] and S. F. Masri[2]

[1]Jet Propulsion Laboratory, California Institute of Technology, M/S 198-235 4800 Oak Grove Drive Pasadena, CA 91109; PH (818) 393-7368; FAX (818) 393-5007; email: mohammad@caltech.edu
[2]Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, 3620 S. Vermont Ave, KAP 210, MC 253, Los Angeles, California 90089-253; PH: (213) 740-0603; FAX: (213) 744-1426; email: masri@usc.edu

ABSTRACT
It is well-recognized that the corrosion of metallic structures has a significant impact (i.e., direct cost of approximately $276 billion per year) on the U.S. economy, including infrastructure, transportation, utilities, production and manufacturing, etc. There is an urgent need to develop more reliable ways to detect corrosion. Image processing techniques have been used extensively to detect corrosion in structures; however, it is essential to evaluate the effect of different parameters on the performance of vision-based corrosion detection systems. This study evaluates the effect of several parameters, including color space, color channels, and sub-image block size on the performance of color wavelet-based texture analysis algorithms for detecting corrosion.

1. INTRODUCTION
Corrosion encompasses an important class of defects in civil infrastructure systems. The application of image-understanding algorithms for corrosion detection is of great interest since they are contactless, nondestructive methods. Furthermore, to segment corrosion-like regions from the rest of an image, both texture and color analysis should be used.

1.1 Background
Furuta et al. (1995) developed a system based on HSV color space and Neural Network (NN) analysis to help inexperienced inspectors in the damage assessment of structural corrosion. Livens et al. (1995) classified corrosion images into two classes by performing a wavelet decomposition of images and using a Learning Vector Quantization (LVQ) network.

Gunatilake et al. (1997) used Daubechies (1992) wavelets of order 6 to detect corroded regions of aircraft skins. A three-level wavelet filter bank was used to decompose the images. A nearest neighbor classifier was used to classify corroded regions from non-corroded ones.

Siegel and Gunatilake (1998) used the YIQ color space, the Battle-Lemaire (BL) (Strang and Nguyen, 1996) wavelet transform filter and a feed-forward NN to classify

image blocks as corrosion with high confidence, or corrosion with low confidence, or corrosion-free.

Choi and Kim (2005) classified several types of corrosion defects. They proposed to use HIS color space, principal component analysis, and varimax approaches to extract appropriate classification features. The Gray Level Co-occurrence Matrix (GLCM) was used for texture feature extraction based on the azimuth difference of points on a surface. The magnification factors of the tested images are between 50 and 500.

Pidaparti (2007) and Palakal et al. (2001) used the Haar wavelet transform to compute the classification features. A K-means classifier was used to classify corroded regions from non-corroded ones.

Jahanshahi et al. (2009) provided a survey and an evaluation of some of the promising vision-based approaches for automatic corrosion detection in civil infrastructure systems. Medeiros et al. (2010) combined the GLCM probabilities and HSI color space statistics to detect corrosion in carbon steel storage tanks and pipelines from a petroleum refinery. It was concluded that the combination of texture and color analysis outperforms individual texture and color analyses.

1.2 Scope
In the current study, the performances of wavelet-based corrosion detection approaches are evaluated by utilizing a variety of color spaces, features, and sub-image window sizes to find the optimum combination of these parameters. Section 2 introduces image-based corrosion detection. Wavelet-based texture analysis and color analysis are discussed in Section 2.1 and 2.2, respectively. The conducted research is described in Section 3. The experimental results and discussion are presented in Section 4. Section 5 includes the summary and conclusions.

2. CORROSION DETECTION
Since the corrosion process can deteriorate the surface of metals, the corroded surface has a different texture and color with respect to the rest of the image. Texture can be regarded as the measurement of smoothness, coarseness, and regularity (Gonzalez and Woods, 1992). Consequently, texture and color analysis approaches are appropriate tools to classify and segment corroded regions. Tuceryan and Jain (1998) and Xie (2008) provide comprehensive reviews of different texture analysis techniques.

2.1 Wavelet-Based Texture Analysis
Discrete Wavelet Transform (DWT) coefficients (Mallat, 1989) are powerful tools for characterizing the appropriate features for texture classification, since the wavelet transform provides a remarkable understanding of the spatial and frequency characteristics of an image (Gonzalez et al., 2004). Moreover, since low frequencies dominate most images, the ability of the wavelet transform to repetitively decompose in low frequencies makes it quite useful for many image analysis tasks (Porter and Canagarajah, 1996).

Figure 1 represents a schematic decomposition procedure of an image by two-dimensional DWT. The input to this system is the initial image, i[th] Approximation; h' and h are low-pass and high-pass decomposition filters, respectively. The words

"Columns" and "Rows" underneath these filters indicate whether the columns or rows of the input should be convolved with the decomposition filter. Since a one-step decomposition of the input with a low-pass and a high-pass filter yields almost twice as much data, a down-sampling (indicated by 2↓) keeps the amount of data almost the same size as the input. The words "Columns" and "Rows" beneath the down-sampling boxes indicate that the down-sampling should take place either over columns or rows (which could be done simply by keeping the even-indexed columns or rows).

The output is the $(i + 1)^{th}$ Approximation, which includes the low frequency characteristics of the input, and is closest in similarity to the input image. There will be three horizontal, vertical, and diagonal details, which include the details of the input in the specified directions. These outputs are the wavelet transform coefficients. Since the $(i + 1)^{th}$ Approximation has the most characteristics of the input image, it can be fed to the decomposition system as an input, and decomposition can take place repeatedly. As the order of decomposition increases, more details will be decomposed from the image.
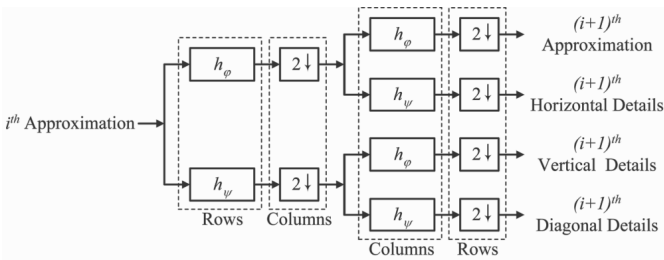
Figure 1: 2D discrete wavelet transform decomposition scheme.

The wavelet filter bank is commonly used for texture classification by first dividing the grayscale image into $m \times m$ pixel blocks (e.g., $m = 8$), followed by computation of the wavelet coefficients for each block. Finally, features are computed for each block based on the wavelet coefficients. These features are used to label blocks as corroded or non-corroded, using a trained classifier. Block-based feature elements lead to a high signal-to-noise ratio and decrease the false detection of corrosion on a surface.

2.2 Color Analysis

Color is another important attribute of digital image-based corrosion detection. Color image segmentation surveys are provided by Skarbek and Koschen (1994) and Cheng et al. (2001). One commonly used approach to integrating the wavelet-based texture analysis and color analysis is to convert the RGB image into an uncorrelated color space. Extraction of the texture features for each decomposed color channel can subsequently be used for classification. The YCbCr, YIQ, and HSI color spaces are considered for color analysis in this study.

The YCbCr color space is usually used for image compression, where Y is the luma component and Cb and Cr are the blue-difference and red-difference chroma components, respectively. The YIQ color space is used by the NTSC color TV

system, where Y represents luminance, and I and Q represent chrominance information. The HSI color space is the closest color space to the human vision system, where H is hue, and S and I are saturation and intensity components, respectively.

In the above color spaces, the color characteristics are independent of brightness, which is suitable for color analysis purposes. One of the objectives of the current study is to find the most appropriate combination of color channels for corrosion detection.

3. CONDUCTED RESEARCH

In this study, each image was first divided into sub-image blocks. Each RGB sub-image was decomposed to less correlated color channels. The YCbCr, YIQ, and HSI color spaces were used for color analysis. Daubechies (1992) wavelets of order 4 were used to establish a three-level wavelet filter bank for each color channel of the sub-images. Ten features for each color channel were computed. The features were presented as the percentage of energy corresponding to the approximation, horizontal, vertical, or diagonal details at each decomposition level.

The energy of each color channel is defined as the sum of the square of all pixel values belonging to that frame, divided by the sum of the square of all pixel values belonging to all decomposed frames of the sub-image. Each feature can be written mathematically as:

$$f_j(i) = \frac{\sum_{(m,n)\in B(i)}(W_j(m,n))^2}{\sum_{j=1}^{10}\sum_{(m,n)\in B(i)}(W_j(m,n))^2} \tag{1}$$

where $f_j(i)$ is the jth feature of the ith block, $W_j(m, n)$ is the wavelet decomposition coefficients of the jth sub-band at (m, n), and $B(i)$ is the ith block. Finally, using the above features, a trained NN is used to classify each block as corroded or non-corroded.

Each image was subdivided into sub-images using a fixed window size (e.g., 8×8). The texture and color analysis approach described above was used to extract features and classify each sub-image. The various color channel combinations shown in Table 1 were used for color and texture analysis to find the optimum combination. Furthermore, in order to find the optimum size of the sub-image window, the following sub-image dimensions were evaluated: 8×8, 16×16, 32×32, 64×64, 96×96, 128×128, 160×160, 192×192, 224×224, and 256×256 (unit = pixel $\times$ pixel).

4. EXPERIMENTAL RESULTS AND DISCUSSION

In order to train and test the above combinations, several images were captured of corroded and non-corroded regions on steel structures. These regions were manually classified and labeled.

Several NNs were trained and evaluated for each scenario. The NNs were all three-layer feed forward networks. The back-propagation method was used to train each network. The input layers consisted of 20 or 30 neurons, depending on the number of selected features. The hidden layer of the NNs was set to have 50 neurons. The output

layers consisted of two neurons. To compare the performance of different approaches, the Area Under Curve (AUC) of the Receiver Operating Characteristic (ROC) curve was computed for each scenario. ROC is a graphical plot of the true positive rate versus the false positive rate for a binary classifier system, as its discrimination threshold is varied (Fawcett, 2006). A classifier with a higher AUC performs better than a classifier with a lower AUC. The AUC can be interpreted as the probability that a randomly selected positive case (e.g., corrosion) will be regarded with greater suspicion than a randomly selected negative case (i.e., corrosion-free). The AUC can take values from 0.0 to 1.0. An ideal classifier has the AUC of 1.0. On the other hand, an AUC of 0.50 means random chance (i.e., no model). It is possible, but rare, to come across AUCs less than 0.50.

Table 1: Different combinations of color features examined in this study

Combination	Color channels used	No. of features
YCbCr	Y, Cb, Cr	30
CbCr	Cb, Cr	20
YIQ	Y, I, Q	30
IQ	I, Q	20
HSI	H, S, I	30
HS	H, S	20

A total number of 1,018 corrosion sub-images and 1,041 non-corroded images were used to train the NNs. From these 2,059 sub-images, 67% of the samples were randomly selected to train the NNs. For validation and initial testing, 28% of the samples were randomly split in half. The remaining samples, that were not exposed to the NNs during training, were used as the testing set to evaluate the performances of the different approaches. To minimize the effect of the training set's distribution on the NNs, the training and testing were repeated ten times for each experiment.

Figure 2 shows the effects of different combinations of color spaces, features, and 'fixed' sub-image windows on the classification performances. Figures 2 (a) and (b) show the mean and standard deviation of the AUCs, respectively, obtained from ten trained classifiers for different color channel combinations. For each of the color combinations, the general trend is that the performance improves as the sub-image window size increases to an optimum size and degrades afterwards. This is consistent with the increase of signal-to-noise ratio when the size of the sub-image window increases until it reaches the optimum size after which the signal-to-noise ratio decreases.

The HS and HSI color channels, which are closest to the human vision system, clearly have the worst performances (as well as the worst variances). This could be due to the nonlinearity of the HSI color space. The high standard deviation for the HSI color space confirms that this color space is not a reliable choice for the described texture analysis approach. Although the maximum mean of AUC is higher for HSI with respect to HS, the latter one has lower standard deviation. This means that when the brightness component is excluded (i.e., the I color channel), a more reliable system is obtained.

(a)

(b)
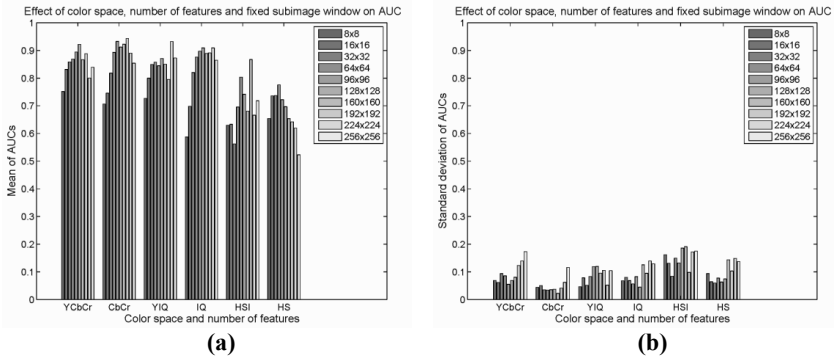
Figure 2: Effect of different color spaces, number of features and different sub-image windows on classification performance: (a) mean of AUCs, (b) standard deviation of AUCs.

From Fig. 2, it is seen that the CbCr color combination has the best performance. The YCbCr, YIQ, and IQ color combinations have close performances. It is worth mentioning that the exclusion of the brightness component in YCbCr and YIQ color spaces (i.e., the exclusion of the Y channel) has led to higher performance when the sub-image window is greater than 64×64 pixels. Furthermore, the reliability of the systems have improved when the brightness component is excluded (i.e., lower standard deviations are obtained). The 192×192 pixel sub-image window utilizing the CbCr color channels has the highest performance.

Tables 2 summarizes various performance indices for the optimum sub-image window sizes. Similarly, Table 3 shows different performance indices for the optimum color channel combination. In these tables, 'accuracy' is the proportion of true classifications in the test set, 'precision' is the proportion of true positive classifications against all positive classifications, 'sensitivity' is the proportion of actual positives that were correctly classified, and 'specificity' is the proportion of negatives that were correctly classified. The latter two quantities, along with the AUC, are insensitive to changes in the class distribution.

5. SUMMARY AND CONCLUSIONS

When using image processing approaches, to segment corrosion-like regions from the rest of the image, both texture and color analysis should be used. Multi-resolution wavelet analysis is a powerful tool to characterize the appropriate features for texture classification. In this study, several experiments were carried out to evaluate the performances of wavelet-based corrosion detection approaches by utilizing a variety of color spaces, features, and sub-image window sizes to find the optimum combination of these parameters.

Different color spaces (including YCbCr, YIQ, and HSI), color channels, and sub-image window sizes were examined to find the optimum color channel combination and sub-image windows sizes to enhance the classification performance. The HSI color space was found to be less consistent, and is not appropriate for using with the proposed multi-resolution wavelet analysis.

Table 2: Different performance indices for the optimum sub-image window sizes

Optimum sub-image		Accuracy (%)		Precision (%)		Sensitivity (%)		Specificity (%)		AUC	
Color	Size	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
YCbCr	128	89.6	5.3	98.3	2.1	83.1	8.3	98.6	1.3	0.92	0.07
CbCr	192	88.6	3.0	98.3	1.6	81.6	3.5	98.3	1.5	0.94	0.04
YIQ	224	90.5	5.0	97.4	3.7	86.3	8.9	95.2	8.1	0.93	0.05
IQ	224	88.1	12.9	91.5	11.6	87.7	14.2	89.4	12.7	0.91	0.14
HSI	192	73.8	12.5	75.3	16.9	84.2	11.9	63.8	21.6	0.87	0.98
HS	64	73.9	10.0	75.7	13.6	80.8	7.2	65.8	18.1	0.78	0.77

Table 3: Different performance indices for the optimum color channel combinations

Optimum sub-image		Accuracy (%)		Precision (%)		Sensitivity (%)		Specificity (%)		AUC	
Color	Size	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
YCbCr	8	70.9	4.9	82.5	7.3	60.8	8.2	83.9	4.8	0.75	0.07
YCbCr	16	79.3	5.0	88.1	4.9	72.0	9.6	87.7	4.8	0.83	0.06
YCbCr	32	83.5	6.6	94.2	3.2	75.2	11.3	94.5	2.5	0.86	0.09
CbCr	64	82.3	4.0	93.5	4.9	73.3	6.1	94.3	2.0	0.89	0.03

The effects of the color channels (i.e., number of features) were also investigated. The results showed that, in most cases, the performance of the corrosion detection system improves when the features obtained from the brightness channel are excluded in the feature selection process. It was also found that optimum systems based on CbCr color channels have the highest performances.

6. ACKNOWLEDGEMENTS

REFERENCES

Bonnin-Pascual, F. and Ortiz, A. (2010). "Combination of weak classifiers for metallic corrosion detection and guided crack location." *Proceedings of the 2010 IEEE Conference on EmergingTechnologies and Factory Automation (ETFA)*. 1 –4.

Buchsbaum,W. H. (1968). *Color TV Servicing*. Prentice Hall Press, Englewood Cliffs, 2nd edition. ISBN 0-13-152389-9.

Cheng, H. D., Jiang, X. H., Sun, Y., and Wang, J. (2001). "Color image segmentation: Advances and prospects." *Pattern Recognition*, 34(12), 2259–2281.

Choi, K. and Kim, S. (2005). "Morphological analysis and classification of types of surface corrosion damage by digital image processing." *Corrosion Science*, 47(1), 1–15.

Chong, K. P., Carino, N. J., and Washer, G. (2003). "Health monitoring of civil infrastructures." *Smart Materials and Structures*, 12(3), 483–493.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia. ISBN 0-89871-274-2.

Fawcett, T. (2006). "An introduction to roc analysis." *Pattern Recogn. Lett.*, 27, 861–874.

Furuta, H., Deguchi, T., and Kushida, M. (1995). "Neural network analysis of structural damage due to corrosion." *Proceedings of ISUMA - NAFIPS '95 The Third International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society*. 109–114.

Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley, Boston, MA. ISBN 0-201-50803-6.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). *Digital Image Processing Using MATLAB*. Prentice Hall, Upper Saddle River, NJ. ISBN 0-130-08519-7.

Gunatilake, P., Siegel, M. W., Jordan, A. G., and Podnar, G. W. (1997). "Image understanding algorithms for remote visual inspection of aircraft surfaces." *Proceedings of SPIE - The InternationalSociety for Optical Engineering*, 3029, 2–13.

Jahanshahi, M. R., Kelly, J. S., Masri, S. F., and Sukhatme, G. S. (2009). "A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures." *Structure and Infrastructure Engineering*, 5(6), 455–486. DOI: 10.1080/15732470801945930.

Livens, S., Scheunders, P., Wouwer, G. V. d., Dyck, D. V., Smets, H., Winkelmans, J., and Bogaerts, W. F. L. (1995). "Classification of corrosion images by wavelet signatures and LVQ networks." *Proceedings of the 6th International Conference on Computer Analysis of Images and Patterns*, CAIP '95, London, UK. Springer-Verlag, 538–543.

Mallat, S. G. (1989). "Multifrequency channel decompositions of images and wavelet models." *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12), 2091–2110.

McCrea, A., Chamberlain, D., and Navon, R. (2002). "Automated inspection and restoration of steel bridges - a critical review of methods and enabling technologies." *Automation in Construction*, 11(4), 351–373.

Medeiros, F. N. S., Ramalho, G. L. B., Bento, M. P., and Medeiros, L. C. L. (2010). "On the evaluation of texture and color features for nondestructive corrosion detection." *EURASIP J. Adv. Signal Process*, 2010, 7:1–7:7.

Pakrashi, V., Schoefs, F., Memet, J. B., and O'Connor, A. (2010). "ROC dependent event isolation method for image processing based assessment of corroded harbour structures." *Structure and Infrastructure Engineering*, 6(3), 365–378.

Motion-Data–driven Unsafe Pose Identification Through Biomechanical Analysis

JoonOh Seo[1], SangUk Han[2], SangHyun Lee[3], and Thomas J. Armstrong[4]

[1]PhD Student, Dept. of Civil and Environmental Engineering, University of Michigan, 2350 Hayward St., Ann Arbor, MI 48109; PH (734) 763-8077; email: junoseo@umich.edu
[2]PhD Candidate, Dept. of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801; PH (217) 244-0691; FAX (217) 265-8039; email: han12@illinois.edu
[3]Assistant Professor, Dept. of Civil and Environmental Engineering, University of Michigan, 2350 Hayward St., Ann Arbor, MI 48109; PH (734) 764-9420; FAX (734) 764-4292; email: shdpm@umich.edu
[4]Professor, Dept. of Industrial and Operations Engineering, University of Michigan, 1205 Beal Ave., Ann Arbor, MI 48109; PH (734) 763-3742; FAX (734) 764-3451; email: tja@umich.edu

ABSTRACT

About 33% of non-fatal occupational injuries and illness are due to work-related musculoskeletal disorders (WMSDs), which had the highest percentage of injuries or illnesses in construction in 2009. Though techniques previously used to prevent WMSDs (e.g., ergonomic rules, checklists based on surveys, and laboratory experiments) provide valuable insights into the prevention of WMSDs, these techniques may not be suitable for measuring the physical demands required for manual ongoing works under real conditions. In an effort to address this issue, we propose a motion capture approach to obtaining a worker's posture information for measuring physical loads on body parts (e.g., shoulder, back). The human postures extracted by motion capture are expressed by rotation angles between body joints, and these angles are then converted to joint angles, which are the inputs for biomechanical analysis. The resulting information contains the internal forces of body parts (e.g., hands and feet), and is capable of identifying allowable strength ranges, thereby helping determine overexertion and awkward postures during a task. As a test case, motion data for ladder climbing is obtained using a motion capture system (e.g., Microsoft Kinect) in a laboratory, and the postures are biomechanically analyzed frame by frame. The results show that the proposed method performs well at computing the physical loads, which promises its great potential to understand the causes of WMSDs during ongoing manual construction work. Further, it can automatically predict potentially hazardous postures on certain body parts, which has otherwise only been achieved by tedious and time-consuming manual investigation.

INTRODUCTION

Construction is associated with a high risk of work-related musculoskeletal disorders (WMSDs) due to the manual-labor–intensive nature of construction works. Twenty-four percent of non-fatal occupational injuries and illnesses in US construction workers were due to WMSDs in 2011 (BLS 2011). WMSDs are a leading cause of lost work days (Meerding et al. 2005), and impact workers' earnings and the contractor's profits (NIOSH 2007). Detecting and reducing exposures to physical risk factors can help to prevent WMSDs (Armstrong et al. 1996) and associated disability (Simoneau et al. 1999). Physical exposures can be assessed by measuring three-dimensional postures and load patterns over time.

A wide range of risk factor assessment methods have been developed: 1) self-reports (e.g., interviews and questionnaires), 2) observational methods, and 3) direct measurements (e.g., sensors) (Li and Buckle 1999; David 2005). Despite their usefulness, limitations exist when these methods are applied in real conditions, due to their subjectivity, the costs of acquisition, the training and time required to undertake assessments and analyze the data, and possible interference with ongoing work (Li and Buckle 1999). To address these limitations, we propose the use of a quasi-static biomechanical analysis of musculoskeletal loads in which the postures are derived from a marker-less motion capture system and the load forces are inferred from empirical models. A Microsoft Kinect motion sensor is used to capture worker postures and movements. This method is demonstrated through ladder climbing case study. We discuss the benefits of the proposed methodology for identifying behavioral risk factors during construction works.

LITERATURE REVIEW

Biomechanical models of the human body have been widely used to understand and reduce the risk of WMSDs in the workplace (Marras and Radwin 2005). Biomechanical models describing the complex musculoskeletal system of the human body (Bean et al. 1988) help to estimate internal forces that can rarely be measured directly (Chaffin et al. 1996). Based on the assumption that the behaviors of the human body follow the laws of Newtonian mechanics, biomechanical models provide a quantitative assessment of the musculoskeletal loadings during occupational tasks, which helps with the identification of hazardous loading conditions on certain body parts (Marras and Radwin, 2005).

Recently, several computerized simulation and analysis tools (e.g., 3D SSPP, Opensim) based on three-dimensional biomechanical modeling have been developed for estimating the musculoskeletal stresses, and provide both proactive and reactive analysis of work tasks. As these tools require motion data to calculate joint angles that are the most important parameters in biomechanical models, previous research efforts utilized motion capture systems in laboratories (Delp et al. 2007, Malineni and King 2010; Symeonidis et al. 2010; Muriti 2005). However, because currently available motion capture systems (e.g., VICON) require a performer to wear a special suit and attach reflective markers to each joint, they are not suitable for collecting

motion data under real conditions due to the possibility of interference with ongoing work (Han et al. 2012).

To overcome this issue, a system based on the Kinect sensor has been applied for ergonomic monitoring during occupational tasks including construction works (Martin et al. 2012; Ray and Teizer 2012). Kinect is a motion-sensing device developed by Microsoft with a maximum sensor range of 4000 mm. The main advantage of the Kinect system is that it does not require any markers or the wearing of a special suit during motion capture, and allows researchers to generate human-skeleton–based motion data using motion capture solutions (e.g., iPi Motion Capture) (Han et al. 2012). Though we have limited use of the Kinect due to relatively short sensing range and light sensitivity in an outdoor environment (Ray and Teizer 2012), it can be a useful tool to collect workers' motion data in limited indoor workplaces (e.g., finishing works in building construction).

METHODS

In this paper, biomechanical analysis is implemented in 3D SSPP, software developed by the Center of Ergonomics at the University of Michigan (Chaffin et al. 2006). With posture data, force parameters, and anthropometry data, workers' motions in 3D space can be simulated in 3D SSPP. Based on the biomechanical simulation, static strength requirements for certain tasks are predicted, including the strength to perform the described job and the spinal compression force using the static model (Center for Ergonomics, University of Michigan 2011) that assumes the effect of acceleration and momentum are negligible.

For posture data, body joint angles that are calculated from motion data can be manually input in 3D SSPP. The magnitude and direction of force data, which is pre-determined in each frame, can be applied for biomechanical analysis.

Based on the analysis results of postures, the body parts that endure forceful exertion can be found. Figure 1 shows an example of the biomechanical analysis result in 3D SSPP. The left three images are the pose with different viewpoints, and the image on the right is the analysis result. The limits in the bar graphs—green to yellow transition, and yellow to red transition—correspond to the NIOSH Action Limit and Maximum Permissible Limit, respectively (NIOSH, 1981). If the bar that represents joint moments in this pose is in the red zone, the body segment has a high risk of getting injured.
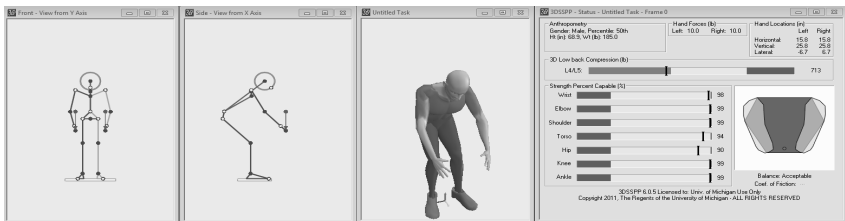
Figure 1. User Interface in 3D SSPP

As a case study, this paper performed the quasi-static biomechanical analysis using the University of Michigan 3D SSPP with motion data from a Kinect sensor, and external force data during ladder climbing. Due to forceful exertion on hands and feet, the repetitive use of a ladder can cause localized fatigue (Kumar 1999), which may contribute to the development of work-related musculoskeletal disorders in the long term. From the analysis, we identified which poses produce excessive internal loads beyond human capacity.

DATA COLLECTION

Motion data for ladder climbing activities was collected using a Kinect as a case study. To collect motion data, the Kinect was used to take depth images frame by frame with a video while the subject repeatedly ascended and descended a portable ladder (25 times). From the depth images, 3D skeletons of the subject were extracted using the iPi Motion Capture solution, and converted into the motion data as a BVH format that defined a hierarchical structure including bone length, initial position and rotation of body parts (See Figure 2).
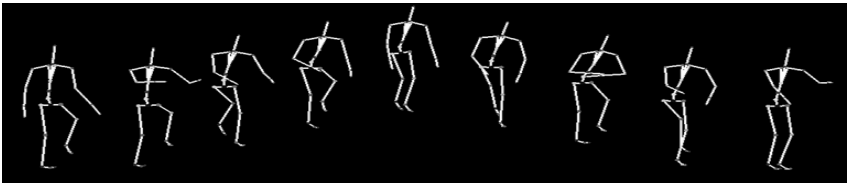
Figure 2. Sequential motions in the BVH format during ladder climbing

While postures in 3D SSPP are defined as the horizontal and vertical angles of body parts, the BVH motion data format expresses sequential postures with rotation angles. This format, however, defines the hierarchical and spatial structure of a human body and stores a 3D position of a root body joint (e.g., a hip); thus, 3D positions of all of the body joints can iteratively be computed from the root joint using translations and rotations (i.e., a transformation matrix). To obtain the body segment angles required for 3D SSPP, the 3D locations (x,y,z coordinates) of body joints were first calculated from the motion data in the BVH format, and then the joint angles were computed based on the vectors of bones in a local coordinate system of the body. For example, the vertical and horizontal angles of upper arms for 3D SSPP (See Figure 3(a)) were calculated from two vectors—a vector from a neck joint to a shoulder joint, and a vector from a shoulder joint to an elbow joint (See Figure 3(b)).
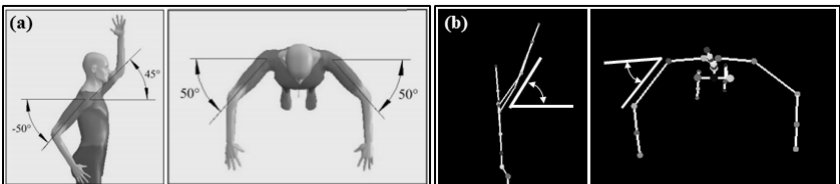
Figure 3. Vertical and horizontal angle of upper arm: (a) in 3D SSPP, (b) in the BVH motion data

External force data during ladder climbing was determined from literature. Armstrong et al. (2008) conducted extensive laboratory experiments to measure hand and foot forces during ladder climbing. From their experiments, they found that the hands exert forces from 28% to 39% of the bodyweight, while the feet support most of the bodyweight (94% to 100%). We assumed that the peak forces are exerted on the hands (39% of the bodyweight, 68 lbs) and feet (100% of the bodyweight, 170 lbs) when the hands and feet are in contact with a ladder. The subject's anthropometric data (72 in, 175lbs) was manually input in 3D SSPP.

RESULTS AND DISUSSION

During ladder climbing, joint moments on body joints (e.g., wrist, elbow, shoulder, hip, knee) and low back stresses are generally under the allowable static maximum (10%–30% of NIOSH's Maximum Permissible Limit (NIOSH, 1981)), except for ankle moment. As shown in Figure 4, the strength percent capability of the right ankle fell in the red area (See A in Figure 4) when the subject raised his left foot to climb the ladder. At this frame, the peak moment exerted in the right ankle was observed, showing a maximum of 1593 lbs, while the maximum permissible limit is 964 lbs (Less than 25% of male workers have this strength.).

Figure 4. A result of biomechanical analysis (one frame)

Figure 5 shows joint moments in the right ankle and the strength limit during a climbing cycle. The average joint moment in the right ankle during one climbing cycle was 600 lbs, while the average strength limit was 1020 lbs. Though the peak moment in the right ankle was repeatedly observed when the subject put his right foot on the rungs of the ladder, excessive joint moments beyond the strength limit were detected only when the subject put his foot on the first rung from the floor.
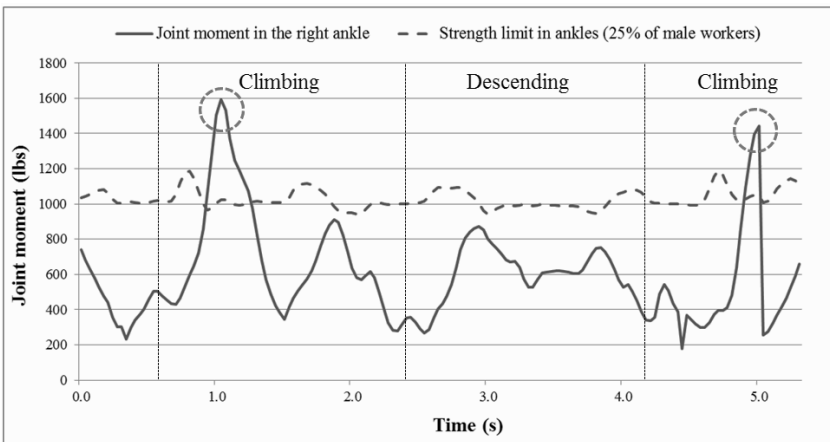
Figure 5. Joint moment in the right ankle and strength limit for 25% of male workers

The result indicates the ergonomic risk inherent in ladder climbing activities. When climbing up, excessive loads can generally be placed on an ankle even for a short interval, and these loads can potentially lead to injuries (e.g., sprains or strains) when repeated, especially for elder people, whose capacity is relatively less than that of young workers. Based on the biomechanical analysis, therefore one can provide proper intervention (e.g., spreading the load by using hands when climbing up) to prevent such injuries.

Taking into account that the magnitude, duration, and repetition of the musculoskeletal loads exerted on the human body are fundamental factors that cause work-related musculoskeletal disorders (Kilbom et al. 1996), quantitative measurement of joint moments exerted on body joints is critical in identifying the risks of these disorders. The proposed approach has great potential as a field-oriented ergonomic evaluation method, illustrating how biomechanical analysis based on workers' postural information can be achieved during construction tasks.

CONCLUSION

The results show that the proposed method performs well at computing the physical loads, which promises a great potential to understand the causes of WMSDs during ongoing manual construction work. Further, the method can automatically predict potentially hazardous postures on certain body parts, which has otherwise only been achieved by tedious and time-consuming manual investigation. The continuous behavior monitoring of biomechanical stresses during construction tasks will provide effective ergonomic interventions (e.g., feedback to workers on their climbing styles), thereby contributing to the prevention of ladder-related injuries.

FUTURE WORK

In order to validate the proposed method, we will compare biomechanical analysis results using the Kinect with those using marker-based motion capture systems (e.g., VICON) that have been generally used in laboratory experiments. If the musculoskeletal stresses obtained from the proposed method are similar to those from laboratory experiments, it can be concluded that the proposed method has a potential as a field-oriented ergonomic evaluation tool.

ACKNOWLEDGEMENT

REFERENCES

Armstrong, T.J., Ashton-Miller, J., Wooley, C., Kemp, J., Young, J. and Kim, H. (2008). Development of Design Interventions for Preventing Falls from Fixed Ladders. CPWR Technical Report.

Armstrong, T., Buckle, P., Fine, L., Hagberg, M., Haring-Sweeney, M., Martin, B., and Viikari-Juntura, E. (1996). "Musculoskeletal Disorders: Work-related Risk Factors and Prevention." International journal of occupational and environmental health, 2(3), 239–246.

Bean, J. C., Chaffin, D. B., & Schultz, A. B. (1988). "Biomechanical model calculation of muscle contraction forces: a double linear programming method." Journal of Biomechanics, 21(1), 59–66.

Bureau of Labor Statistics. (2012). "Nonfatal Occupational Injuries and Illnesses Requiring days away from work, 2011" (http://www.bls.gov/news.release/pdf/osh2.pdf)

Chaffine, D.B., Andersson, G.B. and Martin, B.J. (2006). Occupational Biomechanics, fourth edition, John Wiley & Sons, Inc., Hoboken, New Jersey.

Center for Ergonomics, University of Michigan (2011). *3D Static Strength Prediction Program: User's Manual*. University of Michigan, MI.

David, G. C. (2005). "Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders." Occupational Medicine, 55(3), 190–199.

Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., ... & Thelen, D. G. (2007). "OpenSim: open-source software to create and analyze dynamic simulations of movement." Biomedical Engineering, IEEE Transactions on, 54(11), 1940–1950.

Han, S., Lee, S., & Peña-Mora, F. (2012). "Vision-based motion detection for safety behavior analysis in construction." 2012 Construction Research Congress (CRC), 21–23.

Kilbom, A., Armstrong, T., Buckle, P., Fine, L., Hagberg, M., Haring-Sweeney, M., Martin, B., Punnett, L. and Silverstein, B. (1996). "Musculoskeletal

Disorders: Work-related Risk Factors and Prevention." International Journal of Occupational and Environmental Health. 2(3), 239–246.

Kumar, S. (1999) *Biomechanics in Ergonomics*. Taylor & Francis Inc., Philadelphia, US.

Li, G. and Buckle, P. (1999) Current Techniques for Assessing Physical Exposure to Work-related Musculoskeletal Risks, with Emphasis on Posture-based Methods. Ergonomics 42(5): 674-695

Malineni, S. M., & King, G. W. (2010). "Evaluation of Stepping Task Biomechanics using OpenSim." American Society of Biomechanics Annual Conference, Providence, Rhode Island.

Marras, W.S. and Radwin, R.G. (2005). "Biomechanical modeling." Reviews of Human Factors and Ergonomics, 1(1), 1–88.

Martin, C.C., Burkert, D.C., Choi, K.R., Wieczorek, N.B., McGregor, P.M., R.A. Herrmann and Beling, P.A. (2012). "A real-time ergonomic monitoring system using the microsoft Kinect." IEEE Systems and Information Engineering Design Symposium, Charlottesville, VA, USA, 50–55.

Meerding, W. J., IJzelenberg, W., Koopmanschap, M. A., Severens, J. L., and Burdorf, A. (2005). "Health problems lead to considerable productivity loss at work among workers with high physical load jobs." Journal of clinical epidemiology, 58(5), 517–523.

Muriti, A. J. (2005). A Biomechanical Analysis of Patient Handling Techniques and Equipment in a Remote Setting (Doctoral dissertation, University of New South Wales).

National Institute for Occupational Safety and Health (1981). *Work Practices Guide for Manual Lifting.* Technical report number: 81–122, U.S. Dept. of Health and Human Services (NIOSH),Cincinnati, Ohio.

National Institute for Occupational Safety and Health (NIOSH). (2007). *Simple Solutions: Ergonomics for Construction Workers*, DHHS (NIOSH) Publication No. 2007–122, August 2007.

Occupational Safety and Health Administration (OHSA). (1999). *Preventing Musculoskeletal Disorders in Construction worker,* ACCSH.

Ray, S. J., and Teizer, J. (2012). "Real-time construction worker posture analysis for ergonomics training." Advanced Engineering Informatics.

Simoneau, S., St-Vincent, M. and Chicoine, D. (1996). Work-Related Musculoskeletal Disorders (WMSDs) - A Better Understanding for More Effective Prevention. Technical Guide RG-126-ang, Montréal, IRSST.

Symeonidis, I., Kavadarli, G., Schuller, E., & Peldschus, S. (2010). "Simulation of Biomechanical Experiments in OpenSim." XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010 (pp. 107–110). Springer Berlin Heidelberg.

Information Transformation and Automated Reasoning for Automated Compliance Checking in Construction

J. Zhang[1] and N. M. El-Gohary[2]

[1]Graduate Student, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, IL 61801; PH (217) 607-6006; FAX (217) 265-8039; email: jzhang70@illinois.edu
[2]Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 North Mathews Ave., Urbana, IL 61801; PH (217) 333-6620; FAX (217) 265-8039; email: gohary@illinois.edu

ABSTRACT

This paper presents a new approach for automated compliance checking in the construction domain. The approach utilizes semantic modeling, semantic Natural Language Processing (NLP) techniques (including text classification and information extraction), and logic reasoning to facilitate automated textual regulatory document analysis and processing for extracting requirements from these documents and formalizing these requirements in a computer-processable format. The approach involves developing a set of algorithms and combining them into one computational platform: 1) semantic machine-learning-based algorithms for text classification (TC), 2) hybrid syntactic-semantic rule-based algorithms for information extraction (IE), 3) semantic rule-based algorithms for information transformation (ITr), and 4) logic-based algorithms for compliance reasoning (CR). This paper focuses on presenting our algorithms for ITr. A semantic logic-based representation for construction regulatory requirements is described. Semantic mapping rules and conflict resolution rules for transforming the extracted information into the representation are discussed. Our combined TC, IE and ITr algorithms were tested in extracting and formalizing quantitative requirements in the 2006 International Building Code, achieving 96% and 92% precision and recall, respectively.

INTRODUCTION

Manual regulatory compliance checking of construction projects is costly, time-consuming, and error-prone. Automated compliance checking (ACC) is expected to reduce the time, cost, and errors of compliance checking. Previous research (e.g. Tan *et al*. 2010, Eastman *et al*. 2009, Lau and Law 2004, Garrett and Fenves 1987) and software development efforts (e.g. Solibri 2011) have undoubtedly paved the way for ACC in the architectural, engineering, and

construction (AEC) industry. However, these efforts are limited in their automation and reasoning capabilities; existing ACC systems require manual effort for extracting requirements from textual regulatory documents (e.g. codes) and encoding these requirements in a computer-processable format. To address this gap, the authors are proposing a new approach for ACC (Zhang and El-Gohary 2012). It utilizes semantic modeling, semantic Natural Language Processing (NLP) techniques (including text classification and information extraction), and logic reasoning to facilitate automated textual regulatory document analysis (e.g. code analysis) and processing for extracting requirements from these documents and formalizing these requirements in a computer-processable format. The approach involves developing a set of algorithms and combining them into one computational platform: 1) semantic machine-learning-based algorithms for text classification (TC), 2) hybrid syntactic-semantic rule-based algorithms for information extraction (IE), 3) semantic rule-based algorithms for information transformation (ITr), and 4) logic-based algorithms for compliance reasoning (CR). In this paper, we focus on presenting our algorithms for ITr.

BACKGROUND

Natural language processing (NLP) is a field of artificial intelligence (AI) that is intended to enable computers to analyze and process natural language text or speech in a human-like manner. Examples of NLP techniques include tokenization, part-of-speech (POS) tagging, named entity recognition, and co-reference resolution etc. (Marquez 2000). Information extraction (IE) is a subfield of NLP that aims at extracting targeted information from text sources to fill in pre-defined information templates. In our proposed ACC approach, we utilize NLP techniques because construction codes and regulations are represented in unstructured text format. NLP techniques will facilitate the analysis and processing of these codes and regulations for extraction and formalization of requirements/rules.

PROPOSED AUTOMATED COMPLIANCE CHECKING APPROACH

The authors are proposing a six-phase, iterative approach for extracting requirements from textual regulatory documents and formalizing these requirements in a computer-processable format (Figure 1).
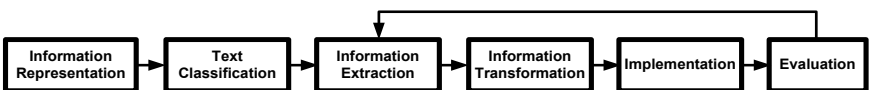

Figure 1. Proposed approach.

Phase 1 – Information representation. We represent the requirements in construction regulations as first order logic-based axioms. Our representation is limited to Horn-Clause-type axioms to facilitate further reasoning using logic programs (logic programming can only represent sentences of the form of a Horn Clause). "Horn clause is a disjunction of literals of which at most one is positive." All horn clauses can be represented in rules that have one or more antecedents (i.e. left-hand sides) that are conjoined (i.e. combined using 'and operator'), and a single consequent (i.e. right-hand side) (Russell and Norvig 2010). Each horn clause represents one requirement. Its right-hand side indicates what this requirement is about. Its left-hand side is consisted of one or more predicates. A predicate is consisted of a predicate symbol and one or more arguments in parenthesis following the predicate symbol. Each predicate organizes information instances corresponding to one or more concepts and/or relations. The horn clauses representation is the target format for our information transformation (ITr) process. The input representation to ITr is the tuple format which results from IE. Each piece in the tuple is an "information element". Each extracted word or phrase recognized corresponding to an "information element" is an "information instance". An example illustrating input and output formats of ITr is shown in Table 1.

Table 1. A transformation example.

Information Extraction (Output: Tuple Format)					
Requirement Sentence	Subject	Compliance Checking Attribute	Comparative Relation	Quantity Value	Quantity Unit/Reference
Courts shall not be less than 3 feet in width.	court	width	not less than	3	feet
Information Transformation (Output: Logic Clause)					
Generated logic clause	compliance_width_of_court(Court) :- width(Width), court(Court), has(Court,Width), greater_than_or_equal(Width,quantity(3,feet)).				

Phase 2 - Text classification. Text classification (TC) aims at recognizing the relevant sentences from a text corpus. Relevant sentences are the sentences that contain the type of information that need to be extracted and transformed into logic clauses. This phase saves unnecessary processing of irrelevant sentences in later phases. It also avoids extraction and transformation errors caused by irrelevant sentences. The presentation of our TC algorithms and results is outside the scope of this paper. For further details on the authors' work in the area of TC, the reader is referred to Salama and El-Gohary 2013.

Phase 3 - Information extraction. Information extraction (IE) aims at recognizing the words and phrases in the relevant sentences that carry target

information, extracting these information, and filling these information into pre-defined information templates. Target information is the needed information for constructing logic clauses that describe requirements in construction regulations. IE consists of feature generation, target information analysis, and development of extraction rules. Both syntactic (i.e. related to syntax and grammar) and semantic (i.e. related to context and meaning) features are used for IE. The presentation of our IE algorithms and results is outside the scope of this paper. For further details on the authors' work in the area of IE, the reader is referred to Zhang and El-Gohary 2012.

Phase 4 - Information transformation. Information transformation (ITr) aims at transforming the extracted information into logic clauses. ITr algorithms are developed using semantic mapping rules, and conflict resolution rules. The semantic mapping rules define how to process the information instances according to their semantic meaning. The semantic meaning of each information instance is defined by the concept or relation it is associated with. (e.g. 'subject' defines the semantic meaning for 'court' in the example in Table 1, i.e. it defines that 'court' is the 'subject' of compliance checking). For example, one semantic mapping rule could be "If both 'subject' and 'attribute' information instances exist for an information tuple, then generate a fresh variable with the 'subject' information instance being the predicate symbol, generate another fresh variable with 'attribute' information instance being the predicate symbol, and a relationship 'has' with the two arguments filled by the two variables". According to this semantic mapping rule, horn clause disjoints court(Court), width(Width), and has(Court,Width) will be generated for the statement "Courts shall not be less than 6 feet in width", since for this requirement sentence, "court" is recognized as 'subject' information instance and "width" is recognized as 'attribute' information instance. Conflict resolution rules resolve conflicts between information elements. For example, one conflict resolution rule could be "The information instance indicating 'comparative relation' should appear before its corresponding information instance indicating 'quantity value', and it should be the nearest one to its 'quantity value'". According to this conflict resolution rule, for the requirement sentence "The openings therein shall be a minimum of 1/8 inch and shall not exceed 1/4 inch", the two 'comparative relation' information instances - 'minimum' and 'not exceed', will be coupled with the correct 'quantity value' information instances - '1/8' and '1/4', respectively.

Phase 5 - Implementation. This phase aims at implementing our algorithms for TC, IE, and ITr into one computational platform. For logic clause representation, we chose the representation of Prolog to facilitate future compliance reasoning (CR). Prolog is an approximate realization of the logic programming computation model on a sequential machine (Sterling and Shapiro 1986). We used the syntax of B-Prolog. B-Prolog is a Prolog system with extensions for programming concurrency, constraints, and interactive graphics. It

has bi-directional interface with C and Java (Zhou 2012). We utilized two types of logic statements in B-Prolog syntax: facts, and rules. A rule has the form: "H :- B1, B2, …, Bn. (n>0)". H, B1, …, Bn are atomic formulas. H is called the head and the right-hand side of ':-' is called the body of the rule. A fact is a special kind of rule whose body is always true (Zhou 2012). To build the ground for quantitative reasoning, we develop a set of built-in rules for our logic clause representation. To prevent non-termination of deduction process, the "cut" operator in Prolog is utilized to remove choice points from alternative clauses to the left of the "cut". TC and IE are implemented using GATE (General Architecture for Text Engineering) tools (Univ. of Sheffield 2011). GATE has a variety of built-in tools for a variety of text processing functions (e.g. tokenization, sentence splitting, POS tagging, gazetteer compiling, morphological analysis, Java Annotation Patterns Engine, etc.). For ITr, the semantic mapping rules and conflict resolution rules are implemented in Python programming language (v3.2.3). The "re" module (i.e. regular expression module) in Python is utilized for pattern matching, so that each extracted information instance could be used for subsequent processing steps based on their information element tags (example tags are shown in Figure 3).

```
<potentialsubject>      Courts</potentialsubject> shall <negation>not</negation> be <less_than>less than</less_than>
<Quantity_Value><Quantity_Value><Quantity_Value>3</Quantity_Value></Quantity_Value></Quantity_Value>
<Quantity_Unit>feet</Quantity_Unit> in <potentialattribute><potentialsubject>width</potentialsubject></potentialattribute>.
```

Figure 3. An example sentence with recognized information element tags.

Phase 6 - Evaluation. This phase aims at evaluating the combined result of TC, IE, and ITr using precision (P), recall (R), and F-measure (F), where P = correct logic clause elements produced / total logic clause elements produced, R = correct logic clause elements produced / total logic clause elements ought to be produced, and F= 2PR/(P+R). A logic clause element is a predicate symbol or a predicate argument for a logic clause. For example, for the predicate court(C), 'court' is a logic clause element, and 'C' is also a logic clause element.

PRELIMINARY EXPERIMENTAL RESULTS AND ANALYSIS

The proposed approach was tested on quantitative requirements in the 2006 International Building Code (ICC 2006). Chapter 12 was randomly selected for testing. A quantitative requirement is a rule that defines the relationship between a quantitative attribute of a subject and a specific quantity. The preliminary experimental results are shown in Table 2.

Table 2. Preliminary experiment results

	Subject	Compliance Checking Attribute	Comparative Relation	Quantity Value	Quantity Unit/Reference	Total
Number of logic clause elements in gold standard	233	163	67	76	132	671
Total number of logic clause elements generated	225	156	69	76	119	645
Number of logic clause elements correctly generated	210	151	63	75	119	618
Precision	0.93	0.97	0.91	0.99	1.00	0.96
Recall	0.90	0.93	0.94	0.99	0.90	0.92
F-Measure	0.92	0.95	0.93	0.99	0.95	0.94

To conduct our experiment, we have developed and used a small-size ontology to assist in the recognition and extraction of construction domain concepts and relations. We used the built-in ontology editor in GATE for ontology development. For TC and IE, we used ANNIE (A Nearly-New Information Extraction System) in GATE for POS tagging, and gazetteer compiling; and we used JAPE (Java Annotation Patterns Engine) transducer for text classification and for writing information extraction rules. When conducting our IE, five information elements were recognized: 'subject', 'compliance checking attribute', 'comparative relation', 'quantity value', and 'quantity unit' or 'quantity reference'. A 'subject' is a 'thing' (e.g. building object, space, etc.) that is subject to a particular regulation or norm. A 'compliance checking attribute' is a specific characteristic of a 'subject' by which its compliance is assessed. A 'comparative relation' is a relation commonly used for comparison such as greater_than_or_equal, less_than_or_equal, or equal_to, etc. A 'quantity value' is the value that quantifies the requirement. A 'quantity unit' is the unit of measure for the 'quantity value'. A 'quantity reference' is a reference to another quantity (a quantity value and its corresponding unit). Ten and 11 semantic mapping rules and conflict resolution rules were developed, respectively.

The preliminary experimental results show more than 90% performance (using measures of P, R, and F) for all information elements. This indicates the potential effectiveness of our approach. Yet, since a 100% performance was not achieved, we conducted an error analysis to identify the sources of error. For example, in one case, the cause for not producing a logic clause element that ought to be produced was the use of uncommon expression structures in the text: in the sentence "…has as a longer side at least 65 percent open and unobstructed", the "quantity reference" was not produced because the adjective "65 percent open and unobstructed" is not a commonly-used expression structure for "quantity reference". For future improvement, to prevent/reduce the detected errors, we propose two ways of adapting and refining our algorithms. First, enhance our ontology-based deductive reasoning to better detect concepts that are implicit in

the text (e.g. the 'compliance checking attribute' 'height' was implicit in the part of sentence "ventilators… at least 3 feet above eave or cornice vents"). Second, add pointer-word resolution rules to avoid missing relations between concepts connected by pointer-words (e.g. the pointer-word 'thereof' in the part of sentence "Any room … in two thirds of the area thereof" connects the 'compliance checking attribute' 'area' with the 'subject' 'room').

CONCLUSION AND FUTURE WORK

In this paper, the authors proposed a six-phase iterative approach for extracting requirements from textual regulatory documents and formalizing these requirements in a computer-processable format. The approach combines our text classification (TC), information extraction (IE), information transformation (ITr), and compliance reasoning (CR) algorithms into one computational platform. The paper focuses on presenting and discussing our information transformation algorithms, which transforms the extracted information, into logic-based representation ready for compliance reasoning. Our approach was tested in extracting and formalizing quantitative requirements in Chapter 12 of the 2006 International Building Code. A precision, recall, and F-measure of 96%, 92%, and 94%, respectively, were achieved. The preliminary experimental results show that our proposed approach is promising. For further improvement, we conducted an error analysis. As part of future/ongoing work, the authors will adapt and refine our algorithms to prevent/reduce the detected errors. Our future work on automated compliance checking will also explore automated IE and ITr from other types of construction documents (e.g. contract specifications).

ACKNOWLEDGEMENT

REFERENCES

Eastman, C., Lee, J., Jeong, Y., and Lee, J. (2009) "Automatic rule-based checking of building designs." *Autom. Constr., 18(8*), 1011-1033.
Garrett, J., and Fenves, S. (1987). "A knowledge-based standard processor for structural component design." *Eng. with Comput.*, 2(4), 219-238.
International Code Council (ICC). (2006). "2006 International Building Code." <http://publicecodes.citation.com/icod/ibc/2006f2/index.htm> (Feb. 05, 2011).
Joshi, A. (1991). "Natural Language Processing." *Science, 253*, 1242-1249.

Lau, G., and Law, K. (2004). "An information infrastructure for comparing accessibility regulations and related information from multiple sources." *Proc., 10th Intl. Conf. Comput. Civ. and Build. Eng.*, Weimar, Germany.

Marquez, L. (2000). "Machine learning and natural language processing." *Proc., "Aprendizaje automatico aplicado al procesamiento del lenguaje natural".*

Russell, S., and Norvig, P. (2010). *Artificial intelligence – a modern approach (third edition),* Pearson Education, Inc., Upper Saddle River, New Jersey.

Salama, D., and El-Gohary, N. (2013). "Automated semantic text classification for supporting automated compliance checking in construction". *J. Comput. Civ. Eng.*, *In Review.*

Solibri.   (2011).   "Solibri   Model   Checker." <http://www.solibri.com/solibri-model-checker.html> (July 15, 2011).

Sterling, L., and Shapiro, E. (1986). *The art of Prolog: advanced programming techniques,* MIT Press, Cambridge, Massachusetts, London, England.

Tan, X., Hammad, A., and Fazio, P. (2010) "Automated code compliance checking for building envelope design." *J. Comput. Civ. Eng.*, 24(2), 203-211.

University of Sheffield. (2011). "General architecture for text engineering." <http://gate.ac.uk/> (Feb. 05, 2011).

Zhang, J., and El-Gohary, N. (2012). "Extraction of construction regulatory requirements from textual documents using natural language processing technieques." *Proc., Comput. Civ. Eng.*, ASCE, Reston, VA, 453-460.

Zhou, N. (2012). "B-Prolog user's manual (version 7.7): Prolog, agent, and constraint   programming."   Afany   Software. <http://www.probp.com/manual/manual.html> (Nov. 19, 2012).

A Transformational Approach to Explicit Stereo Camera Calibration for Improved Euclidean Accuracy of Infrastructure 3D Reconstruction

H. Fathi[1] and I. Brilakis[2]

[1]PhD Candidate, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA; Email: ha_fathi@gatech.edu
[2]Laing O'Rourke Lecturer of Construction Engineering, Department of Engineering, University of Cambridge, UK; Email: ib340@cam.ac.uk

ABSTRACT

The accuracy of the results in stereo image-based 3D reconstruction is very sensitive to the intrinsic and extrinsic camera parameters determined during camera calibration. The existing camera calibration algorithms induce a significant amount of error due to poor estimation accuracies in camera parameters when they are used for long-range scenarios such as mapping civil infrastructure. This leads to unusable results, and may result in the failure of the whole reconstruction process. This paper proposes a novel way to address this problem. Instead of incremental improvements to the accuracy typically induced by new calibration algorithms, the authors hypothesize that a set of multiple calibrations created by videotaping a moving calibration pattern along a specific path can increase overall calibration accuracy. This is achieved by using conventional camera calibration algorithms to perform separate estimations for some predefined distance values. The result, which is a set of camera parameters for different distances, is then uniquely input in the Structure from Motion process to improve the Euclidean accuracy of the reconstruction. The proposed method has been tested on infrastructure scenes and the experimental analyses indicate the improved performance.

INTRODUCTION

Automatic 3D reconstruction of infrastructure from multiple view imagery is considered as an inexpensive alternative to laser-based systems but cannot replace them without acceptable levels of geometrical accuracy (Strecha et al., 2008). Two main issues need to be studied for this purpose: camera calibration and dense multi-view geometry (Strecha et al., 2008). The scope of this paper is to focus on the first issue because it is the first step in the 3D reconstruction pipeline. Camera calibration is the process of determining a set of parameters that describe the mapping between 3D world coordinates and 2D image coordinates. The existing methods are divided into two categories: a) explicit calibration (i.e., conventional approach) and b) self-calibration. Methods in the first category estimate the calibration parameters by establishing correspondences between reference points on an object with known 3D dimensions and their projection on the image. On the other hand, self-calibration automatically provides necessary parameters through the geometrical constraints in images, but is less accurate than the explicit methods (Furukawa and Ponce, 2009). This paper aims to focus on the explicit approach because accuracy is typically the main concern in the 3D reconstruction of infrastructure.

Most of the existing image-based reconstruction methods in computer vision use a single camera for image acquisition. This imposes a constraint on the generated results because, by using a single camera, a scene can only be reconstructed up to an unknown scale factor (Pollefeys et al., 2008). This limitation is of great importance especially in infrastructure applications that require spatial data collection in the Euclidean space. The use of a calibrated stereo camera set up eliminates this problem but at the cost of additional steps for camera calibration and more sensitivity of the results to the calibration parameters (Peng, 2011; Xu et al., 2012).

This paper initially aims to demonstrate that existing explicit stereo camera calibration procedures can potentially induce a significant amount of error when the results are used in long-range applications. The impact of using calibration information determined from conventional methods on the Euclidean accuracy of 3D points with different range values is studied for this purpose. As the main contribution of this paper, a new approach is then presented to address this problem. This approach does not provide new mathematical relationships for estimating the necessary parameters; instead, conventional calibration algorithms are used in a novel procedure. It provides a set of multiple calibration parameters that are estimated for some predefined distance values between the calibration object and the camera set. The sets of calibration parameters are ultimately used in the bundle adjustment step of the Structure from Motion (SfM) process to achieve optimum X,Y,Z coordinates for 3D points. This allows maintaining the well-known benefits of the conventional algorithms while improving the Euclidean accuracy of the reconstruction; preliminary results from a case study (a building façade) show an average of ±3.1cm (25%) improvement in the accuracy of 3D points.

BACKGROUND

The first step in calibrating a camera is to define a camera model. In computer vision, most practical cameras are represented by a pinhole camera model (Fig. 1). In this model, each point in the world space $(X,Y,Z)^T$ is projected by a straight line into the image plane, through the camera center C. The calibration

Figure 1: Pinhole camera model

parameters in this model are categorized into intrinsic and extrinsic. Intrinsic parameters represent internal geometry and optical characteristics of the lens and include focal length (f_x, f_y), principal point (u,v), and distortion coefficients $(k_1, k_2, p_1, p_2, k_3)$. On the other hand, the camera position (t) and orientation (R) in the 3D space comprise extrinsic parameters. If image points are represented by homogeneous vectors, a 3D point is projected on the image plane as
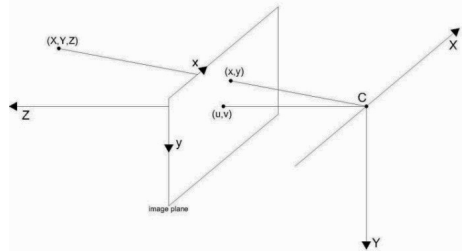
$$\begin{pmatrix} x'' \\ y'' \\ w \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + t \tag{1}$$

$$x_d = x''/w \quad , \quad y_d = y''/w \quad , \quad r^2 = x_d^2 + y_d^2 \tag{2}$$

$$x' = x_d(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 x_d y_d + p_2(r^2 + 2x_d^2) \tag{3}$$

$$y' = y_d(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y_d^2) + 2p_2 x_d y_d \tag{4}$$

$$x = f_x x' + u \quad , \quad y = f_y y' + v \tag{5}$$

In the explicit camera calibration, a set of images is captured from different views of a checkerboard or an object with known dimensions. Then, correspondences between 3D coordinates of corner points and their 2D image projections are established. Eqs. 1-5 are finally used in a non-linear optimization problem to estimate the unknown camera parameters. This general process has been used in most of the existing camera calibration methods such as Zhang (2000), Wang et al. (2008), and Furukawa and Ponce (2009). Although these methods have been mainly proposed for a single camera, they can be expanded to the stereo camera calibration scenario. In this case, the relative position between two cameras (R_0, t_0) needs to be found beside the intrinsic and extrinsic parameters of each camera. The geometric relationship between the left (R_l, t_l) and right (R_r, t_r) cameras can be expressed as follows

$$R_0 = R_r R_l^{-1} \quad , \quad t_0 = t_r - R_r R_l^{-1} t_l \tag{6}$$

These conventional methods have been successfully used in close-range 3D reconstruction applications with spatial accuracies that rival laser scanning (Seitz et al., 2006); however, the same level of accuracy has not been achieved in calibrated stereo image-based long-range applications even using high resolution cameras (Fathi and Brilakis, 2011). It is known that the accuracy of results is very sensitive to the intrinsic and extrinsic stereo camera calibration parameters (House and Nickels, 2006; Geiger et al., 2011). This may be justified by the point that in a calibrated stereo image-based application, the estimated parameters are kept constant in all subsequent steps of a SfM process and errors can accumulate. Moreover, conventional methods model the lens distortion using five coefficients that represent radial and tangential distortions. These coefficients have the highest impact on the Euclidean accuracy of 3D points compared to other parameters, especially in long-range scenarios which a small deviation between the estimated and actual distortion can result in significant amounts of errors (Peng, 2011). It has been shown that this effect is expected to be minimal for a 3D point, when the Z coordinate of the point is very close to the average distance used between the camera and calibration object in the calibration step (Peng, 2011). These indicate a need for further research on both the calibration and dense 3D reconstruction processes.

The research objective of this paper is to a) test whether a proposed multi-step camera calibration procedure can enhance the Euclidean accuracy of the results and b) study the impact of using stereo camera calibration information that is acquired from the conventional methods on the Euclidean accuracy of 3D points in different range values. The key research question that will be answered is: how can we use the known information about sensitivity of Euclidean accuracy of 3D points with respect

to calibration parameters and design a camera calibration procedure that is capable of providing higher Euclidean accuracies?

METHODOLOGY

A set of two video cameras capable of streaming raw video data and a pair of fixed focal length lenses are used to address the objectives; this is required to avoid the change of focal length and information loss during image compression. An appropriate baseline distance is also selected based on the typical range values that are encountered in infrastructure applications and the stereo camera system is set up. Once the sensor system is ready, the following parameters should not change while collecting the necessary data: video resolution, focal length, and relative position of the two cameras. A checkerboard with an appropriate number of black and white squares in two perpendicular directions is also required for the calibration process. The number of squares and their dimensions are selected according to the scene.

Inspired by the results of previous studies (House and Nickels, 2006; Strecha et al., 2008; Peng, 2011; Xu et al., 2012) and considering the abovementioned control variable settings and delimitations, the authors hypothesize that the following procedure for stereo camera calibration can enhance the final Euclidean accuracy of 3D points. If the distance between the sensor system and the calibration board is denoted by D, the conventional explicit stereo camera calibration procedure is repeated i times ($i = 1,...n$) for different D values. At each repetition, a different value is selected for D (e.g., $D_i = 10m$) and while it is fixed, a set of stereo video streams are collected (Fig. 2). The camera system and the board move in a way that D_i does not change significantly but the board can be seen from different angles. The collected data is then used as the input in a conventional calibration algorithm to find the required parameters. The result corresponding to D_i is saved and the process is repeated for another D. The outcome is a multiple set of calibration parameters $\{P_i \,|\, i = 1,...,n\}$, each corresponding to a specific D. It is necessary to mention that while videotaping, it is preferred to move the camera set such that the calibration board appears at different areas of video frames. It is known that if the calibration pattern only appears at the central part of video frames, the estimations will behave poorly at peripheral areas (Zhang, 2000).

The sensor system is then used to collect stereo videos from a desired infrastructure scene. In the processing step (i.e., SfM), the results of the proposed multi-step stereo camera calibration procedure are used. First, the average of each calibration parameter is calculated from $\{P_i \,|\, i = 1,...,n\}$. In the visual triangulation
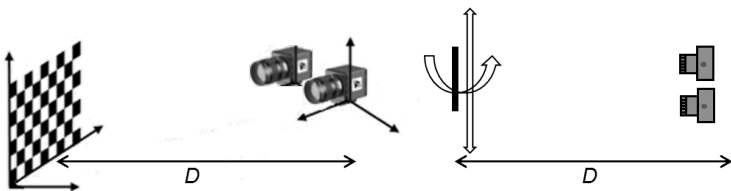

Figure 2: Data collection schematic for a D value (left: side view, right: top-down view)

step, these average values are used to find an initial estimation of 3D coordinates of points. For point j in k-th stereo view (p_{jk}), the P_i that has the closest D to the Z coordinate of p_{jk} is found. The corresponding calibration parameters are then tied to p_{jk} and the 3D coordinates are recalculated. This new information is inputted in the bundle adjustment process to achieve the final estimation for 3D points.

DESIGN OF EXPERIMENTS

In order to study the impact of using conventional stereo camera calibration procedures on the accuracy of 3D coordinates of points versus the proposed one, a well-textured planar scene is needed to be selected for two reasons: a) it allows controlling the Z coordinate of 3D points in the desired range by simply changing the distance between a stereo camera system and the planar face; and b) well-textured regions provide the opportunity to detect and match enough number of feature points in stereo views such that the results can be statistically significant.

For camera calibration, six sets of stereo video streams are captured from the board under different conditions. In the first set which will be used for testing conventional procedures, the distance between the camera and the board changes in the range of $5m \leq D_1 \leq 15m$ while capturing the videos. Captured video frames should cover different views and angles of the board while the camera moves smoothly toward and/or away from the board. The next five sets are needed to test the proposed stereo camera calibration procedure. In these sets, the distance between the camera system and the board is fixed to $D_2 = 5m$, $D_3 = 10m$, $D_4 = 15m$, $D_5 = 20m$, and $D_6 = 25m$, respectively. These limits have been selected according to the typical range values that we encounter in building applications. The sensor system is also used to collect stereo videos from the planar scene while the distance of the camera to the planar scene changes from $5m \leq D \leq 25m$. This data is a control variable and will be used for 3D reconstruction of the scene in two scenarios: a) using conventional calibration procedures (parameters acquired from the 1st set of calibration videos); and b) using the proposed multi-step calibration procedure (multiple set of parameters acquired from the 2nd to 6th sets of calibration videos).

The performance of the proposed calibration procedure is assessed based on the following metrics: a) spatial accuracy of the initial estimation for 3D coordinates of points with different range values; and b) spatial accuracy of a dense 3D point cloud. For the first metric, stereo frames corresponding to $D = \{5, 10, 15, 20, 25m\}$ are extracted from the planar scene video to detect and match feature points. Calibration parameters from the conventional and proposed procedures are then used to estimate 3D coordinates of feature points from left and right views of stereo frames. Spatial distance between pairs of feature points is then calculated for each case and compared to the ground truth data that is acquired using total station surveying. For the second metric, calibration parameters from the conventional and proposed procedures are used separately in a dense 3D reconstruction package and the spatial accuracy of the results is evaluated. The sample size at all experiments is considered to be 50 which corresponds to 90% confidence level and ±10% confidence interval.

IMPLEMENTATION AND RESULTS

A prototype was created using Microsoft Visual C# to implement and test the proposed multi-step calibration procedure. The C# platform provides a base to connect to any number of cameras with real-time responsiveness. OpenCV (Intel® Open Source and free C++ Computer Vision Library) was selected as the main image processing library. Two high resolution Flea2 cameras were used to capture stereo video streams. The baseline distance was approximately 30cm and the video resolution was 5MP with a frame rate of 7.5 fps. A calibration board with a pattern of 13×14 squares each with a dimension of 60mm was also built.

An automatic stereo camera calibration program was developed using the functions available in OpenCV. The user is able to run the program while videotaping a calibration pattern. The program is real-time responsive and automatically detects the pattern in every frame. Once the pattern is successfully detected in a pair of stereo frames, corners are automatically detected and matched between the two views (Fig. 3). This process continues until enough number of views, that cover different angles of the pattern, are captured (typically 30-50). Then, the stereo calibration function is invoked to calculate the required parameters.

Figure 3: Automatically detected and matched calibration board corners

A set of experiments were performed according to the details presented in the design of experiments section. A planar façade with a brick pattern (resulting in many feature points) was selected for data collection. The abovementioned camera system and calibration board were used to capture the six sets of required data for calibration. Using the automatic calibration program, 50 stereo frames were extracted in each case and the calibration parameters were calculated. Then, another set of stereo video streams were captured from the façade while the distance between the camera system and the façade was changing in the range of $5m \leq D \leq 25m$.

In the first step of the designed performance evaluation process, stereo frames corresponding to $D = \{5,10,15,20,25m\}$ were extracted from the façade video and 3D coordinates of feature points were calculated using the sets of estimated calibration parameters. Spatial distance between pairs of 3D feature points was then compared to the ground truth data. Table 1 illustrates the average error at each scenario (sample size of 50). The preliminary results indicate that a more accurate initial estimation can be done for a point at a range of Z using the calibration parameters that correspond to $D \approx Z$. This is because of better estimation for lens distortion.

Table 1. Average spatial distance error for different calibration scenarios

Calibration Scenario	Average spatial distance error (cm)				
	$Z \approx 5m$	$Z \approx 10m$	$Z \approx 15m$	$Z \approx 20m$	$Z \approx 25m$
$D = 5m$	**±2.6**	±4.8	±9.8	±17.7	±23.8
$D = 10m$	±2.9	**±4.3**	±8.5	±15.9	±20.2
$D = 15m$	±3.5	±5	**±6.3**	±14.5	±19.6
$D = 20m$	±4.2	±6.1	±9	**±11.3**	±18.4
$D = 25m$	±5.1	±7.5	±12.7	±15.2	**±15**
$5m \leq D \leq 25m$	±3.2	±4.9	±10.1	±15.6	±19.2

For the second step of the performance evaluation process, two dense 3D point clouds of the façade (using all the frames in the façade video) were generated. The results of the conventional camera calibration procedure were used to generate the first point cloud. The second one was generated using the calibration parameters acquired from the proposed procedure (Fig. 4). 50 pairs of points were selected at each case and the spatial distance was compared to the ground truth data. The average error in the first point cloud was ±12.6cm while this average error was ±9.5cm in the second point cloud. This shows an average of 3.1cm (25%) improvement in the accuracy of results because of using the proposed multi-step stereo camera calibration procedure. It is necessary to mention that this accuracy can be further improved by modifying the multi-view geometry process which is the target of our future work.
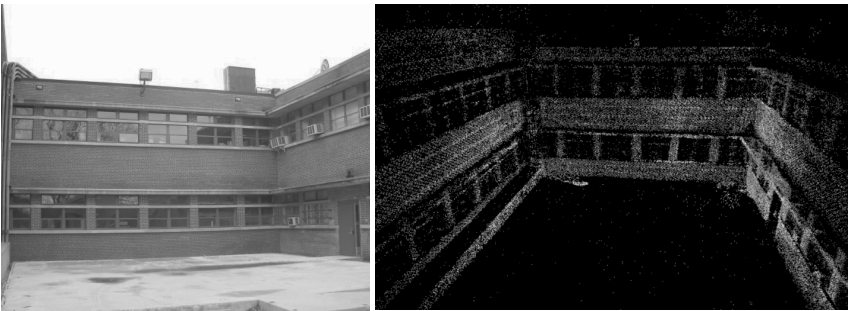
Figure 4: Dense 3D point cloud of the façade

SUMMARY AND CONCLUSION

Accurate 3D reconstruction of infrastructure from multiple-view imagery can provide the construction industry with an inexpensive alternative to the laser-based surveying techniques. In the case of using a calibrated stereo camera system, it has been shown that the accuracy of final results is very sensitive to the calibration parameters; while the highest sensitivity corresponds to the distortion coefficients. Due to this sensitivity, the existing stereo camera calibration algorithms only provide accurate results when they are used in close-range applications.

This paper presented a novel multi-step stereo camera calibration procedure to alleviate the abovementioned problem. The goal was to enhance the Euclidean accuracy of 3D reconstruction in long-range scenarios. The proposed procedure uses

a set of discrete values to represent the distance between the sensor system and the calibration board (D). For each D, a set of stereo video streams are collected while the distance between the camera and the board is fixed to D. Conventional stereo camera calibration algorithms are then used to calculate calibration parameters for the given D. Repeating this process for all the values results in a multiple set of parameters each corresponding to a specific D. The sets of calibration parameters are then used in the SfM process with the following assumption: for each 3D point, the set of calibration parameters that have the closest D value to the point's Z coordinate are used. Preliminary results demonstrated that this procedure is capable of reducing the spatial measurement errors by 25%. As mentioned before, camera calibration and dense multi-view geometry are key issues regarding the capability to achieve an acceptable level of accuracy. Therefore, even more accurate results are expected to be acquired once the second issue (i.e., dense multi-view geometry) is also studied in future works. It is anticipated that solving these two issues can reduce the errors to practical levels (less than 2-3cm).

REFERENCES

Fathi, H., and Brilakis, I. (2011). "Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features", *Advanced Engineering Informatics*, 25(4), 760-770.

Furukawa, Y., and Ponce, J. (2009). "Accurate camera calibration from multi-view stereo and bundle adjustment", *International Journal of Computer Vision*, 84(3), 257-268.

Geiger, A., Zeigler, J., and Stiller, C. (2011). "StereoScan: dense 3D reconstruction in real-time", *Proc. IEEE Intelligent Vehicles Symposium*, 963-968.

House, B., and Nickels, K. (2006). "Increased automation in stereo camera calibration techniques", *Systemics, Cybernetics and Informatics*, 4(4), 48-51.

Peng, J. (2011). "Comparison of three dimensional measurement accuracy using stereo vision", *Thesis, University of Regina*.

Pollefeys, M., Nister, D., Frahm, J., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, R., Welch, G., and Towles, H. (2008). "Detailed real-time urban 3D reconstruction from video", *International Journal of Computer Vision*, 78 (2-3), 143-167.

Seitz, S., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). "A comparison and evaluation of multi-view stereo reconstruction algorithms", *Proc. IEEE Conference on CVPR*, 519-526.

Strecha, C., Von Hansen, W., Van Gool, L., Fua, P., and Thoennessen, U. (2008). "On benchmarking camera calibration and multi-view stereo for high resolution imagery", *Proc. IEEE Conference on CVPR*, 1-8.

Wang, J., Shi, F., Zhang, J., and Liu, Y. (2008). "A new calibration model of camera lens distortion", *Pattern Recognition*, 41(2), 607-615.

Xu, G., Chen, L., and Gao, F. (2012). "Study on binocular stereo camera calibration method", *Proc. International Conference on IASP*, 133-137.

Zhang, Z. (2000). "A flexible new technique for camera calibration", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.

Multisensor Data Fusion for Determining Hallway Blockages in a Building During Evacuation

G. GUVEN[1], E. ERGEN[2], B. OZBAS[3], M.A. ERBERIK[4], O. KURC[5], M.T. BIRGONUL[6]

[1]Doctoral Student, Civil Engineering Department, Istanbul Technical University, Istanbul, 34469, TURKEY; PH (90) 212 285 7399; email: gursans.guven@itu.edu.tr
[2]Assistant Professor, Civil Engineering Department, Istanbul Technical University, Istanbul, 34469, TURKEY; PH (90) 212 285 6912; email: esin.ergen@itu.edu.tr
[3]Graduate Student, Civil Engineering Department, Istanbul Technical University, Istanbul, 34469, TURKEY; PH (90) 212 285 7399; email: ozbas@itu.edu.tr
[4]Associate Professor, Civil Engineering Department, Middle East Technical University, Ankara, TURKEY; PH (90) 312 210 2443; email: altug@metu.edu.tr
[5]Assistant Professor, Civil Engineering Department, Middle East Technical University, Ankara, TURKEY; PH (90) 312 210 2447; email: kurc@metu.edu.tr
[6]Professor, Civil Engineering Department, Middle East Technical University, Ankara, TURKEY; PH (90) 312 210 2427; email: birgonul@metu.edu.tr

ABSTRACT

When a building is under the threat of a disaster/an emergency, it is important to rapidly retrieve up-to-date indoor navigation conditions by identifying the hallways that are blocked. This paper investigates the contribution of data from each sensor that is deployed inside a building in accurately determining the current blockage conditions at hallways. Sensor performances were assessed both individually and by fusing multisensor data at the building element-level. The final goal is to create and visualize evacuation and navigation paths for the first responders and the occupants during a disaster/an emergency. A prototype was developed to integrate Building Information Model (BIM) with hallway blockage information retrieved from multiple sensors and to provide guidance in a building based on the building's current status using graph theory principles. This paper specifically discusses if less sensors could be used to get similar accuracy results in detecting hallway blockages based on the individual and fused data from multiple sensors.

INTRODUCTION

Obtaining timely and accurate information related to a damaged building is critical when a building is under the threat of a disaster/an emergency. This information will be used to guide building occupants safely and quickly out of a building. Previous experience shows that even panicked occupants of a damaged building tend to comply with the authorized visual and/or audible guidance, as observed in the case of 9/11 terrorist attack in which the victims followed the directions of 911 operators' (National Commission 2004). Also, the emergency

response teams require such critical information for entrance to the building and for safe evacuation of the building. To effectively navigate in the building, information regarding the hallways that are blocked following a disaster/an emergency is needed. In the current practice, it is difficult to access this information at the disaster/emergency area and the information that is available is incomplete and inadequate for an effective indoor navigation (Evans et al. 2005; Son and Pena Mora 2006; Ergen and Seyis 2008). Thus, there is a need for an approach that can guide the occupants during evacuation and assist the emergency response teams by providing reliable and accurate post-disaster/emergency information of a building.

The study presented in this paper is part of a research study that develops a framework to assist the first responders and the occupants during a disaster/an emergency by creating and visualizing evacuation and navigation paths. The developed prototype integrates BIM with hallway blockage information retrieved from multiple sensors and provides indoor guidance based on the current status of the building using graph theory principles. Different types of sensors that are deployed inside the building provide information regarding the current status of building elements. Data from these sensors will be fused to understand the current blockage conditions. Using multiple sensors ensure that if some of the sensors are damaged during the disaster/emergency, remaining sensors can still provide some information. However, heavily damaged buildings, where most of the systems are not working properly, are out of the scope of the developed framework. More information on the developed framework can be found in Guven et al. (2012). In this paper, contribution of each sensor data in accurately determining hallway blockage status is determined and possibility of using less number of sensors is investigated. The results of this study were used for choosing appropriate type and number of sensors for data fusion.

BACKGROUND RESEARCH

Different sensors were fused to obtain the blockage conditions of a building in case of a disaster. Previous sensor fusion studies related to buildings mostly aimed for occupancy detection in office buildings and utilized different kinds of sensors such as magnetic loop counters, optical tripwires, optical motion detectors, high-dynamic range cameras, lightning control and CO_2 sensors (Dodier et al. 2006; Hutchins et al. 2007; Sarkar et al. 2008; Erickson et al. 2009; Lam et al. 2009; Meyn et al. 2009). Model-based methods, such as Bayesian probability theory, Markov chain, Monte Carlo sampling and multivariate Gaussian model, were used for fusing multisensor data. However, these methods are costly and complex (Goulet et al. 2010).

Sensor fusion studies in construction domain focused on variety of purposes such as locating and tracking materials and components at site (Ergen et al. 2007; Grau and Caldas 2009; Razavi and Haas 2010), and monitoring labor productivity (Navon 2005) by integrating data obtained from Radio Frequency Identification (RFID) and Global Positioning System (GPS) technologies. Researchers also integrated Real-time Location Sensing (RTLS) technologies with Physiological Status Monitors (PSMs) for analyzing ergonomic and safe work conditions of workers (Cheng et al. 2012). Moreover, the position data of construction equipment and personnel obtained from a 3D range camera were fused for improvement of personnel safety (Teizer et al. 2007). Common approach in these studies was to develop ad-hoc

data fusion solutions for the targeted problems. In addition, there have been attempts to formalize frameworks that can automatically identify and execute a sequence of steps on the basis of a given data fusion query (Pradhan and Akinci 2012).

TEST SETUP

To test the accuracy of multiple sensors in detecting hallway blockage and to develop an approach for data fusion, two scaled experimental models (i.e., wide and narrow) that represent hallway of a pilot building were constructed. The depth of the wide hallway model was 4.0 m; the height and width of the hallway were 1.20 m and 1.70 m, respectively. The narrow hallway model, with a width of 0.90 m, was constructed to examine the effect of the width of a hallway on the damage and the resulting blockage. The wider hallway model was made up of three suspended ceilings, four masonry walls and two bookshelves which serve as non-structural objects (Figure 1).
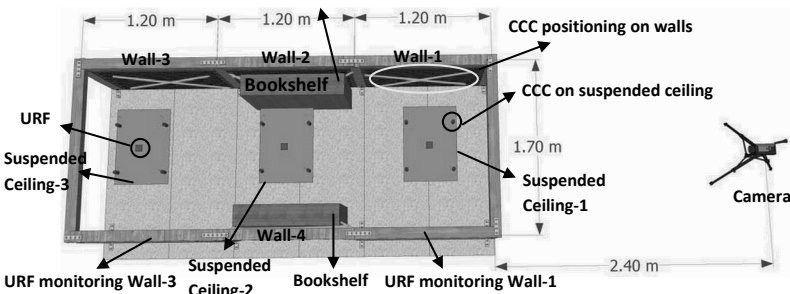
Figure 1. Positions of the building elements and the dimensions of the wider hallway model on a 3D sketch.

The structural frame (i.e., columns and beams) of the model was constructed of wood. Non-structural walls were made of perforated brick units or cardboards and suspended ceilings were made of cardboards. Finally, bookshelves that were made of wooden planks were used to represent the items that can block the passage during a disaster/an emergency. The hallway was divided into three spans, each of which includes four columns and beams. In the topological map of the hallway, each span of the hallway was represented by a node. Blockage information obtained from the related sensors was evaluated separately for each node. The narrow hallway model did not include suspended ceilings, but similar to the wider model, it consisted of walls and bookshelves. More information about the models, experiments and damage scenarios can be found in Ergin (2012).

In total, 189 experiments were conducted (i.e., 168 on wider hallway and 21 on narrow hallway). 29 sensors were installed at building elements placed at various locations in the hallway model and a video camera recorder was placed at one end of the hallway to monitor the hallway (Table 1). In addition to the camera, two types of sensors were used: (1) Closed Cable Circuit (CCC) and (2) Ultrasonic Range Finder (URF).

Table 1. Number and location of sensors used in the wider hallway tests

Element	# of CCC sensors	# of URF sensors
Bookshelf 1	2	-
Bookshelf 2	2	-
Suspended Ceiling 1	4	1
Suspended Ceiling 2	4	1
Suspended Ceiling 3	4	1
Wall 1	2	1
Wall 2	2	-
Wall 3	2	1
Wall 4	2	-
Total	24	5

Figure 2 shows the sensors attached on each type of element. Bookshelves were only monitored with two CCC sensors. The CCCs on the bookshelves were connecting the upper left and right sides of the bookshelves to the columns (upper left connection in Fig. 2a). Suspended ceilings were monitored by four CCCs that were connecting the suspended ceiling to the hallway ceiling at its corners. Also, a URF sensor was attached to the ceiling to monitor the change in the distance between the suspended ceiling and the hallway's ceiling (Fig 2b). Wall-2 and Wall-4 were monitored with CCCs that were positioned crosswise (Fig 2c). Wall-1 and Wall-3 were additionally monitored with one URF sensor, each of which was attached to a cardboard that represent a shear wall and placed on the opposite side of the wall (Fig. 2c).
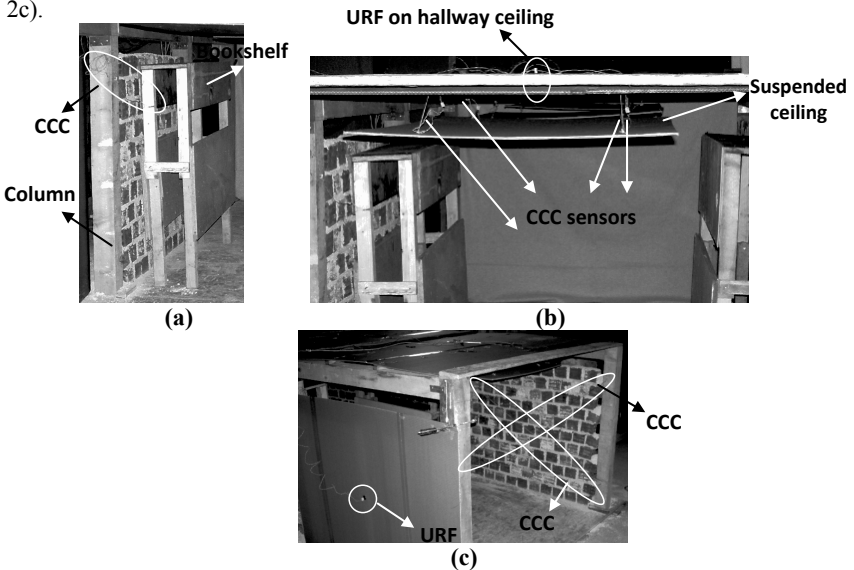

Figure 2. (a) CCC on bookshelves, (b) CCCs and URF on suspended ceiling, (c) CCC on walls and URF (only on Wall-1 and Wall-3).

The camera was placed 2.40 m away from the hallway to monitor all the nodes in the hallway at once (as shown in Fig. 1). During the tests, building elements were manually damaged to simulate the impact of a disaster (e.g., earthquake) on the hallway. Different combinations of possible damage conditions for each node were performed and the resulting blockage levels were determined. For example, bookshelves were pushed down, suspended ceiling connections were detached manually, and walls were demolished. In each test, different scenarios were applied, for example, in once case only ceilings were damaged; in another case, ceilings and bookshelves were damaged at the same time.

During the tests, the data related to the hallway blockage was obtained from the sensors and the camera in real-time. The raw data obtained from CCC sensors were in binary format. Damage given to the element causes the CCC to detach from the element that it is attached and "1" is received from CCC, meaning that the monitored element is damaged. On the other hand, URF sensors provided data in terms of distance (i.e., meters). A threshold percentage value (i.e., %95 for suspended ceilings, %115 for walls) was determined to convert the raw URF data into binary format. When the difference between the initial distance and the distance measured after the damage exceeds the threshold value, it was assumed that the element was damaged and its position was changed.

DATA FUSION

Different levels of blockage conditions were represented using a five-level damage classification: (A) no damage, node is open for passing through; (B) low level of damage, passage through the hallway is not prevented; (C) moderate level of damage, passage is closed for disabled people; (D) high level of damage, people with no disabilities can pass through the hallway by applying force, and (E) very high damage, node completely blocked for passage. During the tests, it was observed that each building element gets damaged in several ways, and this causes different levels of blockage. For example, when all the connections of the suspended ceilings were detached and the bookshelf was damaged at the same time, class D damage occurs on the hallway. But when only the bookshelf was damaged, it results in class C damage. Thus, to determine the blockage levels in hallways, it is not enough to understand whether the building element was damaged or not, but also it is necessary to know how the building element was damaged and whether it is blocking the passage or not.

There are seven suspended ceiling positions that can be observed after a disaster/an emergency and affect the blockage level in different ways: (1) no connections detached, (2) all connections detached, (3) one connection detached, (4) crosswise connections detached, (5) connections parallel to the camera view are detached, (6) connections perpendicular to the camera view are detached, (7) three connections detached. Bookshelves can be in three situations following a disaster/an emergency: (1) bookshelf stands up against the wall, (2) bookshelf fell down, and (3) two bookshelves fell down. Finally, walls can be observed in two situations: (1) wall is not damaged, (2) wall is damaged. Since having both one and two walls cause the hallway to have class E blockage, two damaged walls were not considered.

In the first step, every CCC and URF sensor data was evaluated individually) to determine the accuracy of detecting the given positions of the related hallway

elements. Only four CCCs on a suspended ceiling was not individually assessed since the position of a suspended ceiling can only be determined by combing data from four CCCs attached. In total, five evaluations were performed to asses the accuracy of: (1) each of two CCC sensors in detecting the situation of walls, (2) each of two CCC sensors in detecting the position of bookshelves, (3) all four CCC sensors in detecting the condition of suspended ceilings, (4) the URF sensor on detecting the position of the suspended ceiling, and (5) the URF sensor on detecting the situation of the wall. The results show that, each CCC sensor was able to accurately determine the situation of a wall with 96% accuracy. Similarly, position of bookshelves could be determined by one CCC with 97% accuracy. In 146 tests out of 168 tests where suspended ceilings were damaged, four CCCs were able to correctly detect the position of suspended ceilings (i.e., 87% success rate). The URF can detect the change in the position of a suspended ceiling whenever the ceiling was damaged with a 93% accuracy rate. However, the URF alone cannot identify the predetermined seven positions of a suspended ceiling, and this highlights that URF data should be combined with another sensor data to determine blockage condition. Finally, the accuracy rate of URF in monitoring the walls was also 90%.

In the second step, individual sensor data were fused to determine a building element's status based on some rules driven from the analyses of the experiment results. Element-level fusion of sensors were performed by integrating data from (1) two CCCs on a bookshelf, (2) two CCCs on a wall, (3) one URF and four CCCs on a suspended ceiling, and (4) one URF and two CCCs on a wall (Table 2).

Table 2. Element-level fusion combinations

Bookshelf		Wall			Suspended Ceiling				
CCC-1	CCC-2	CCC-1	CCC-2	URF	CCC-1	CCC-2	CCC-3	CCC-4	URF
X	X								
		X	X						
					X	X	X	X	X
		X	X	X					

When the results from two CCCs attached on a bookshelf were analyzed, the accuracy was %99, whereas individual accuracy of a CCC was 97%. Similarly, the success rate of two CCCs attached on a bookshelf was %98, while the success rate of each CCC was 96%. Therefore, using only one CCC sensor on a wall (Fig 2a) or a bookshelf (Fig 2c) would be sufficient to determine the damage on these building elements instead of using two CCCs.

Although the URF can detect the change in the position of a suspended ceiling whenever the ceiling was damaged with %93 accuracy, data from URF could not determine if the suspended ceiling blocks the hallway (i.e., how the suspended ceiling got damaged). While the URF can only indicate that there is damage, only CCC sensor readings can show the exact type of damage that the suspended ceiling had (Fig. 2b). Therefore, it was concluded that the URF has no significant contribution and that four CCC sensors can be sufficient to determine blockage caused by suspended ceilings. Similarly, the results showed that fusing URF readings with the

readings of CCCs does not improve the blockage assessment results for the walls (Fig. 2c). Hence, using only CCC sensors can be sufficient for walls, as well.

The unit price of a URF sensor is approximately $50 while CCC sensors are less expensive, with a unit price below $1 per meter. Since the system proposes to use the cameras that exist in the building, the total cost of equipping a building with the proposed system can be calculated based on the number of URF and CCC sensors that are needed for a specific building. Thus, the total system cost will depend on the characteristics of the building.

CONCLUSION

In this paper, the contribution of each sensor data in accurately determining hallway blockage status is determined and possibility of using less number of sensors to determine blockage in hallways was investigated. Several sensors were deployed inside a scaled building model to obtain information regarding the blockage levels of the hallways for guiding occupants and first responders during evacuation. The individual accuracy of each sensor was determined and element-level fusion of different and/or same types of sensor data was performed. It was observed that both CCC and URF sensors had high success rates in correctly determining the changes in the elements' conditions. However, URF sensors alone are not capable of providing information regarding how the element is blocking the passage way. Thus, element-level fusion results show that using only CCC sensors were sufficient for understanding element conditions. The results provided an insight on the type and number of sensors that can be used with specific building elements in real life and how the sensors can be deployed on such elements. As future work, sensor fusion will be performed for examining the accuracy of sensors on determining the blockage conditions of nodes (i.e., node-level fusion). Also, contribution of the camera data on the accuracy of the results will be investigated. For improving the test setup, more detailed experiments can be performed by testing the scaled models on a shake table that simulates an earthquake.

ACKNOWLEDGEMENTS

REFERENCES

Cheng, T., Migliaccio, G.C., Teizer, J., Gatti, U.C. (2012). "Data Fusion of Real-time Location Sensing (RTLS) and Physiological Status Monitoring (PSM) for Ergonomics Analysis of Construction Workers", J. Comput. Civ. Eng., ASCE, doi:10.1061/(ASCE)CP.1943-5487.0000222

Dodier, R.H., Henze, G. P., Tiller, D.K., Guo, X. (2006). "Building occupancy detection through sensor belief networks", Energy and Buildings, 38.

Ergin, T. (2012). "A methodology for real-time sensor-based blockage assessment of building structures during earthquakes", MSc thesis, METU, Ankara, Turkey.

Ergen, E., Akinci, B., Sacks, R. (2007). "Tracking and locating components in a precast storage yard utilizing radio frequency identification technology and GPS" Autom. Constr., 16(3), 354–367.

Ergen, E., Seyis, S. (2008)." Utilizing Advanced Storage Technologies For Providing Local Information to Rescue Teams Following an Earthquake", 5th Intl. Conf. on Innovation in Arch., Eng., June 2008, Antalya, Turkey.

Erickson, V.L., Kamthe, Y., Lin, A., Brahme, R., (2009). "Energy efficient building environment control strategies using real-time occupancy measurements". BuildSys '09, Berkeley, USA, pp.19-24.

Evans, D.D. (2003). "First responders problems and solutions: tactical information", Technology in Society, 25(4), 523–528.

Grau, D., Caldas, C.H. (2009). "A methodology for automating the identification and localization of construction components on industrial projects." J. Comput. Civ. Eng., 23(1), 3–13.

Goulet, J.A., Kripakaran, P., Smith, I.F.C. (2010). "Multimodel Structural Performance Monitoring." J. Structural Engineering, 136(10), 1309-1318.

Guven, G., Ergen, E., Erberik, M.A., Kurc, O., Birgonul, M.T. (2012). "Providing guidance for evacuation during emergency based on a real-time damage and vulnerability assessment of facilities", ASCE International Conference on Computing in Civil Engineering, 17-20 June, FL, USA.

Hutchins, J., Ihler A., Smyth, P., (2007). "Modeling count data from multiple sensors: a building occupancy model". CAMPSAP 2007, 241-244.

Lam, K.H., Höynck, M., Dong, B., Andrews, B., Chiou, Y., Zhang, R., Benitez, D., Choi, J., (2009). "Occupancy detection through an extensive environmental sensor network in an open-plan office building". IBPSA, Glasgow, Scotland.

Meyn, S., Surana, A., Lin, Y., Oggianu, S.M., (2009). "A sensor-utility-network method for estimation of occupancy in buildings" Proc. IEEE Conference on Decision and Control & 28th Chinese Control Conf., Shanghai, P.R. China.

Navon, R. (2005). "Automated project performance control of construction projects", Autom. Constr., 14(4), 467–476.

National Commission on Terrorist Attacks (2004). *The 9/11 Commission Report*. <http://govinfo.library.unt.edu/911/report/911Report.pdf>, (Dec, 2012).

Pradhan, A., Akinci, B., (2012). "Planning-Based Approach for Fusing Data from Multiple Sources for Construction Productivity Monitoring", J. Comput. Civ. Eng., ASCE, 26(4), 530-540.

Razavi, S. N., Haas, C. (2010). "Multisensor data fusion for on-site materials tracking in construction." Autom. Constr., 19(8), 1037–1046.

Sarkar, A., Fairchild, M. Salvaggio, C. (2008). "Integrated daylight harvesting and occupancy detection using digital imaging", doi:10.1117/12.765961.

Son, J., Pena-Mora, F. (2006). "Improvement of collaboration among first responders including civil engineers during disaster response in urban areas", 11th ICCBE, June 14-16, Montreal, Canada.

Teizer, J., Caldas, C.H., Haas, C.T. (2007). "Real-time three dimensional occupancy grid modeling for the detection and tracking of construction resources." J. Constr. Eng. Manage., 133(11), 880–888.

Linear Cellular Automation Method for Predicting Dynamic Instability Mode of Single-layer Shell

Guangchun Zhou[1], Ming Zhang[2], Yu Zhang[3] and Yaqub M Rafiq[4]

[1,2,3]School of Civil Engineering, Harbin Institute of Technology, China

[4]University of Plymouth, Drake Circus, UK, PL48AA

[1]gzhou@hit.edu.cn; [2]zming@hit.edu.cn; [3]zhangyuhit@hit.edu.cn; [4]M.Rafiq@plymouth.ac.uk

ABSTRACT

This paper presents a linear cellular automaton (LCA) method to predict the dynamic instability (DI) mode of large-span single-layer shell subjected to ground motions. The LCA model of the shell obtains the state values of cells/nodes through its finite element analysis (FEA). Meanwhile, the concepts of nodal domain and nodal domain similarity are proposed. Then, similar nodal domains between two shells are matched through the proposed criterion. Finally, the DI mode of object shell is mapped using the criterion for projecting the formative values of base shell to similar nodal domains in object shell. Case studies show that the LCA method could be used to predict the DI mode of object shell. Consequently, this study would explore an application of LCA technique and FEA data in shells. Also, the LCA method costs much less time than FEA to calculate the DI mode of shell.

Keywords: shell, linear cellular automaton, dynamic instability, nodal domain, criterion

INTRODUCTION

The latticed shell structures develop rapidly all over the world due to their aesthetic qualities, large space and sound mechanical performance. An outstanding example is that the latticed shells successfully survived during the 1995 destructive Kobe earthquake in Japan, which demonstrated excellent seismic performance clearly (Kunieda and Kitamura 1996). This further aroused researchers' interest in the analysis of shell structures under dynamic loads.

Early thin shell structures could be an empirical and conceptual construction product from human's bionic practice. Then, remarkable performance of this type of structures has led to the construction of many modern large span shell structures around the world. Recently, the development of construction materials, advanced analytical methods and modern construction techniques seems to encourage more use of this type of structures, particular in the seismic areas. However, the following issues still need to address:

(1) A hot issue is to identify and model the failure mechanism of large-span single-layer shells under the strong seismic loading (Kumagai and Ogawa 2003, Shen et al. 2005, Zhi et al. 2010). Commonly, structural instability is believed to be the main cause

of failure in shell structures subjected to seismic excitations. But, structural instability is so complex in its mechanism that it is difficult to determine the corresponding load and mode (Du et al. 2011).

(2) The FEA is widely used in structural analysis and design at present. But, the FEA computation of shell structures undergoing whole seismic process costs great time, particularly, the post-processing of huge calculation data. Hence, alternative convenient and highly efficient methods are expected to address the computing cost.

(3) As Kunieda and Kitamura (1996) indicated, the analysis of shell structure is still needed to verify whether the existing shell structures were designed too strongly or not.

(4) The existing design codes around the world also lack sufficient guidance to support a confident and reliable design approach.

(5) Results obtained from the FEA simulations and tests of shells ought to be further investigated or modeled to discover the new knowledge.

In view of the issues mentioned above, an exploration of both modeling and predicting methods is needed to further new analytical techniques, to apply the test and numerical results and to reduce the computing time of shell structures. Zhou et al. (2006) successfully tried this expectation in mapping the cracking pattern of masonry wall panels using cellular automata, which provides a reference basis for predicting the DI modes of shells.

This paper conducts a LCA modeling of shell structure, because the shell structure itself is a latticed form and the structural nodes are similar to cells in a LCA model. Thus, it is appropriate to develop the concepts of nodal domain and nodal domain similarity, as well as the criteria for matching nodal domain similarity for shell structures. Finally, the LCA method is proposed to predict the DI modes of object shells structures based on the DI mode of a based shell.

CONCEPTS PROPOSED IN THE LCA METHOD

Based on the cellular automata developed by Neuman (1966), Margolus and Toffoli (1987), the concepts for the LCA method are introduced as follows:

Nodal domain: A node and its six neighbouring nodes are defined as a nodal domain, as shown in Fig. 1.

(a) the ith node and its   (b) the kth node and its
neighbourhoods in the object shell   neighbourhoods in the base shell
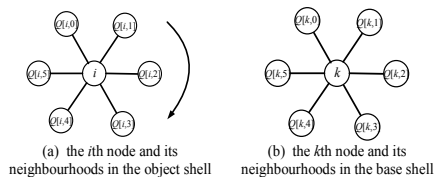Fig. 1 Nodes in the nodal domain

Nodal state value: The normalized FEA displacements of all nodes are defined as nodal state values, for the shell under a unit load. In other words, nodal state values are set as the state values of any cells in the LCA model of the shell.

Nodal domain similarity: It is defined as nodal domain similarity that two nodes with their neighbourhoods have the same or similar node states values (Zhou, 2006, Zhang et al. 2010).

DI mode: the DI mode is composed of all the normalized nodal displacements once the shell becomes instable.

Base shell: A typical shell whose DI mode is known from the FEA numerical simulation result or testing data. The normalized DI mode provides a basis or reference for predicting the DI mode of object shell.

Object shell: A shell whose DI mode needs to be predicted. Here, the object shell's FEA numerical simulation results are available to verify the LCA method.

DATABASE PREPARATION

Verification of the FEA program. The testing and FEA models are shown in Fig. 2. Details of testing process are not shown due to the page limits. The lab test of the shell model was conducted in order to validate the FEA program. Table 1 lists the first and second natural frequencies (f_1, f_2) obtained from the testing and the FEA simulation. Table 1 indicates that the FEA frequencies are very close to the lab testing ones.

(a) The testing model (b) The FEA model
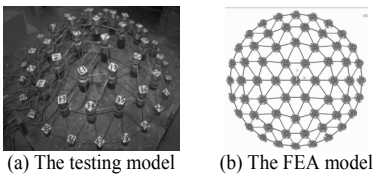
Fig.2 The testing & FEA Shell Models

Table 1. The testing and FEA frequencies

	Test model (Hz)	FEA model (Hz)	Error （%）
f_1	22.16	22.18	0.09
f_2	24.29	24.12	0.69

Database. After validating the FEA program, the database for shells shown in Fig. 3 is established as a basic component for the LCA method. The database consists of two parts: (1) the normalized FEA results of the base and object shells under unit loading case; (2) the normalized FEA (or testing) nodal displacements once the base shell becomes instability.
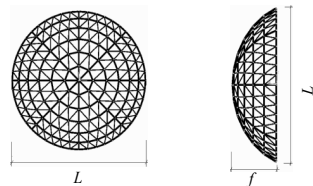
Fig.3 The FEA model of the shell

THE PROPOSED LCA METHOD

The LCA model of the shell. The shell can be constructed as a LCA model. The LCA is a subclass of cellular automata consisting of a lattice of sites on a cylinder evolving according to a linear local interaction rule (Erica 1988). The shell nodes could be considered as the cells of the LCA model. The time and spatial expression of the LCA model is set to correspond to the sections (1) and (2) below:

(1) The linear relationship exists between seismic intensity and displacement response of the shell before its plastic state. The nodal displacement of the shell with viscous damping under harmonic excitations (Chopra 2009) can be expressed as Eq. (1):

$$\{u(t)\} = \frac{P_0}{K_n} \sum_{n=1}^{N} A_n \{\phi\}_n \sin(\omega t + \varphi_n) \tag{1}$$

where A_n is a constant; P_0 is loading amplitude; K_n is generalized stiffness. If the excitation is the same seismic wave with different amplitudes, P_{01} and P_{02}, the ratio of the nodal displacements corresponding to two different seismic amplitudes can be given as Eq. (2)

$$\frac{u_1(t)}{u_2(t)} = \frac{P_{01}}{P_{02}} \sum_{n=1}^{N} \frac{\sin(\omega t + \varphi_1)}{\sin(\omega t + \varphi_2)}; \text{ if } \varphi_1 = \varphi_3, \text{ then } \frac{u_1(t)}{u_2(t)} = \frac{P_{01}}{P_{02}} \tag{2}$$

Eq. (2) indicates that all nodal displacements are linearly related to the amplitude of ground motion, coinciding with the linear property in the LCA mode of the shell. Thus, a node in the state I(i+1) can be updated by itself and its neighbourhoods in the state I(i)

$$\{S\}_{I(i+1)}^{n \times 1} = [T]^{n \times n} \{S\}_{I(i)}^{n \times 1} \tag{3}$$

where $\{S\} = \{a_0, a_1, \cdots, a_n\}^{\mathrm{T}}$ is the state vector of the LCA cells; $[T]$ is a constant transformation matrix; I(i+1) and I(i) are two adjacent seismic intensities.

(2) The spatial property of the LCA model has existed in the database in which the fine FEA simulation has given out an implicit relationship between a cell and its neighbourhoods.

Actually, Eq. (3) also embodies three LCA properties, parallel, locality and homogeneity. For the property of parallel, the state values of individual cells can be updated synchronously. For the property of locality, the state value of a cell depends on the state values of itself and neighbouring cells. For the property of homogeneity, the same rules can be applied to each cell.

The LCA state value. For the state value of a cell in the LCA state mode, Eq. (4) is given as:

$$S_i = u_i^s / \max(u_i^s), \quad i = 1, 2, \cdots, N \tag{4}$$

where S_i is the ith nodal (or cell) state value; u_i^s is the ith nodal displacement under the unit load calculated by the FEA; $\max(u_i^s)$ is the maximum among all the nodal displacements; N is the nodal number.

The criterion for matching nodal domain similarity. For all the nodal domains in the base shell or the object shell, the criterion for matching nodal domain similarity is given by Eq. (5)

$$E_i\left(k,e,j\right)=\min_{k=1}^{N_{\text{base}}}\left(\left\{\min_{j=0}^{5}\left(\left[\left|S[i]-S[k]\right|+\sum_{e=j}^{5+j}\left(S\left[Q(i,e-j)\right]-S\left[Q\left(k,\text{rem}\left(e,6\right)\right)\right]\right)\right]_j\right)\right\}_k\right) \qquad (5)$$

where $S[i]$ refers to the ith nodal state values; Q refers to the neighbour nodes of a node; N_{object} and N_{base} are the nodal numbers of the object and base shells, respectively; rem (e, 6) is the remainder of the integer variables e and a constant integer 6 (the nodal number for a nodal domain).

$E_i\left(k,e,j\right)$ represents two processes: (1) Find the minimum of the state value errors corresponding to six different orientations of two nodal domains, as shown in Fig. 1; (2) Find the minimum of the state value errors between one nodal domain in the object shell and all the nodal domains in the base shell. Thus, each nodal domain on the object shell can find its similar nodal domain on the base domain.

The criterion for projecting state values of DI mode. When the instability of the base shell occurs, the normalized nodal displacements $u_{\text{base}}^{\text{f}}$ are called as the formative values of DI mode and all of them compose of the DI mode of the shell. Thus, a criterion can be given to project the formative values of the base shell to the similar nodes in the object shell, as shown by Eq. (6)

$$u_{\text{object}}^{\text{f}}\left(i\right)\Leftarrow u_{\text{base}}^{\text{f}}\left(k\right),\quad i=1,2,\cdots,N_{\text{object}},\quad k=1,2,\cdots,N_{\text{base}} \qquad (6)$$

where $u_{\text{object}}^{\text{f}}\left(i\right)$ is the formative value of the ith nodal on the object shell. $u_{\text{base}}^{\text{f}}\left(k\right)$ is the formative value of the kth nodal on the base shell. Eq. (6) means that the formative values $u_{\text{base}}^{\text{f}}\left(i\right)$ on the base shell are assigned to the similar nodes in the object shell. Furthermore, all the formative values $u_{\text{object}}^{\text{f}}\left(i\right)$ of the nodes on the object shell are normalized again by Eq. (7)

$$f^{\text{object}}\left(i\right)=u_{\text{object}}^{\text{f}}\left(i\right)\Big/\max_{j=1}^{N_{\text{object}}}\left[u_{\text{object}}^{\text{f}}\left(j\right)\right] \qquad (7)$$

Finally, all the normalized formative values f^{object} compose of the DI mode of the object shell.

The LCA method. Based on the LCA model of the shell, the method can be established to predict the DI mode of the object shell. Fig.4 shows the procedure of the LCA method: Step 1. Build the LCA models for both object and base shells. Then, the state values in the LCA models are obtained from the database; Step 2. Find the nodal similarity domains between the base and object shells, using the state values obtained and the rule given by Eq. (5); Step 3. Construct the DI mode of the base shell utilizing the displacement values at the shell's instability state; Step 4. Map the DI mode of the object shell through a criterion given by Eq. (6).
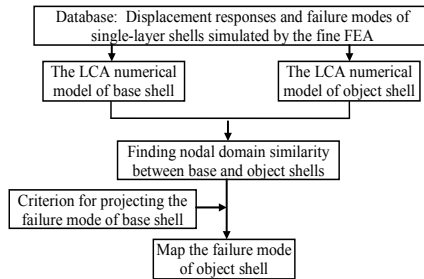
Fig.4 The LCA method for predicting the DI mode of single-layer shell

SIMILARITY BETWEEN DI MODES

Eq. (8) below is introduced to evaluate the similarity between the DI modes of the base and object shells. In other words, Eq. (8) is used to verify the accuracy of the LCA method:

$$\eta = 1 - \frac{1}{E}\sum_{i=1}^{n}\left|f_i^{\text{T}} - f_i^{\text{P}}\right|, \qquad i = 1, 2, \cdots, n \tag{8}$$

where η is the similarity; $E = \sum_{i=1}^{n}\left|f_i^{\text{T}} + f_i^{\text{P}}\right|$; f_i^{T} refers to the ith nodal formative value of

the object shell from the FEA or testing data; f_i^{P} represents the ith nodal formative value predicted by the LCA method; n is the number of the nodes in the object shell.

NUMERICAL EXAMPLES

The following examples are presented to validate the LCA method. Two cases are considered as: Case 1. The DI mode of a base shell is used to predict the DI mode of the object shells with different configurations, for the same dynamic loading case; Case 2. The DI mode of the base shell under a seismic frequency is used to predict the DI mode of the same shell under the different seismic frequencies. The results are shown in Fig. 5. It can be seen from Fig. 5: (1) The DI modes of the object shells predicted by the LCA method are very close to those simulated by the fine FEA models; (2) The average similarity in these examples is 95.82%; (3) The normalized displacements of both base and object shells in their DI states are also close to each other. More examples are not shown here due to space constraints; (4) The LCA method spends much less computing time than that taken by the fine FEA simulation, as verified by the Big-O notation (Cormen et al. 2009)

$$\text{The LCA: } T(n) = O(n^2); \text{ The FEA: } T(n_e) = n_1 \times n_2 \times n_3 \times n_4 \times n_5 \times n_6 = O((n_e)^6) \tag{9}$$

where n is the nodal number, n_1 is the iterative number of mechanical constitutive relations for materials; n_2 is the integral points on the cross section of the elements; n_3 is the number of elements; n_4 is the maximal substeps in each step of the input seismic

wave; n_5 is the input number of the seismic wave; n_6 is the change number of the amplitude of the seismic wave.
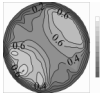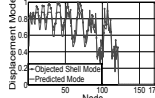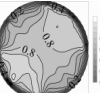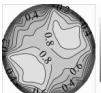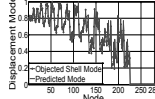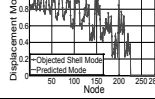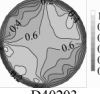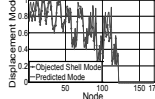
Fine FEA DI mode of base shell	Fine FEA DI mode of object shell	DI mode of object shell predicted by LCA method	Displacements of base and object shells	Similarity
The base and object shells have different ratios of rise-span.				
D40205 Computing time: 168h	D40203 Computing time: 160h	Computing time: 5sec		94.69%
The base and object shells have different spans.				
D50203	D60203			96.35%
The base and object shells have different member cross-sections.				
D60203a	D60203b			98.20%
The base and object shells undergo different frequency harmonic ground motions.				
D40203 (Loading frequency 5Hz)	D40203 (Loading frequency 2Hz)			94.04%
Note	The shell is notated based on its span (m), roof weight (×10kg/m²) and ratio of span-rise. For D40203a, D: dome shell, 40: the shell span (40m), 20: the roof weight (200kg/m²), 3: the ratio of the span-rise, a: the cross section of the member. CPU: i5-2000@ 2.8GHz; Memory: 1.48GHz, 3.24GB			

Fig. 5 The DI modes of shells and comparison of mode similarity

CONCLUSION

(1) The LCA technique is applicable to model the configuration of large-span single-layer shell. In the LCA method, the new concepts of nodal domain and nodal domain similarity, and the criteria for matching nodal domain similarity and projecting the formative values of the base shell could imply the relationship between the nodal domain similarity and the local working behaviour of the shell.

(2) The LCA method further explores the application of the existing FEA (or testing) results and a new way to model the output of the FEA simulation.

(3) The LCA method can save significant time in the prediction of the DI mode of the object shell when compared with the fine FEA method.

It should be noted that the present LCA method is limited to predict the DI displacement mode of the single-layer shell; for the dynamic strength failure mode, it

needs to further study.

ACKNOWLEDGEMENTS

REFERENCES

Chopra, A.K. (2009). "Dynamics of structures theory and applications to earthquake engineering." *Pearson Education*.

Du, W.F., Yu, F.D., and Zhou, Z.Y. (2011). "Dynamic stability analysis of K8 single layer latticed shell structures suffered from earthquakes." *Applied Mechanics and Materials*, 94-96: 52-56.

Erica, J. (1988). "Linear Cellular Automata and Recurring Sequences in Finite Fields." *Communications in Mathematical Physics*, 119(1): 13-28.

Kumagai, T. and Ogawa, T. (2003). "Dynamic buckling behaviour of single layer latticed domes subjected to horizontal step wave." *Journal of the International Association for Shell and Spatial Structures*, 44(3): 167-174.

Kunieda, H. and Kitamura, K. (1996). "Response analysis of cylindrical roof shells subject to Kobe earthquake by mathematically analytic method. " *Proceedings of Asia-Pacific Conference on Shell and Spatial Structures*, Beijing: 712-719.

Margolus, N. and Toffoli, T. (1987). "Cellular Automata Machines: An Object Environment for Modeling." *MIT Press,* Cambridge*,* Mass.

Neumann, J.V. (1966). "The Theory of Self-Reproducing Automata." *University of Illinois Press Champaign*, IL, USA.

Shen, S.Z. and Zhi, X.D. (2005). "Failure mechanism of reticular shells subjected to dynamic actions."*China Civil Engineering Journal*, 38(1):1-20. (in Chinese).

Cormen H.T., Leiserson E.C., Rivest L.R. and Stein C. (2009). "Introduction to Algorithms" Third Edition, *The MIT Press*.

Zhang, Y., Zhou, G.C., Xiong, Y. and Rafiq, M.Y. (2010). "Techniques for predicting cracking pattern of masonry wallet using artificial neural networks and cellular automata." *Journal of Computing in Civil Engineering*, ASCE, 24(2): 161-172.

Zhi, X.D., Fan, F. and Shen, S.Z. (2010). "Elasto-plastic instability of single-layer reticulated shells under dynamic actions." *Thin-Walled Structures*, 48(10-11):837-845.

Zhou, G.C., Rafiq, M.Y., Bugmann, G. and Easterbrook, D.J. (2006). "Cellular automata model for predicting the failure pattern of laterally loaded masonry panels." *Journal of Computing in Civil Engineering*, ASCE, 20(6): 400-409.

Applying Magnetic Charged System Search Algorithm to the Construction Project Planning Problems

Mehdi Tavakolan[1], Mohammad Ali Motie Share[2]

[1] Assistant Professor, School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran

[2] Graduate Student, School of Civil Engineering, College of Engineering, University of Tehran, Tehran, Iran

Abstract

This paper presents Magnetic Charged System Search Algorithm (MCSS) of a new Meta-Heuristic Algorithm (MHA) to solve complex, Time-Cost-Optimization (TCO) problems in construction project planning. Project planners face complicated multivariate TCO problems that require time-cost tradeoff analysis to make models closer to the actual projects. Our approach is applied to solve one optimization problem, which is found in the construction project planning literature. It is shown that our proposed approach is superior than existing optimization algorithms in finding better project schedule solutions with less total project cost and shorter total project duration. The results also show that comparing to other existing methods, the presented approach provides more accurate and complete Pareto front for TCO problems in construction project planning.

Keywords: Optimization, Magnetic Charged System Search Algorithm, Construction Management

Introduction

Project control plays an important role for contractors in scheduling and cost analysis. The time and cost of projects are related to each other and considered in Time-Cost trade-off problems. From the researchers' point of view, developing highly efficient and robust algorithms to solve highly complex Time-Cost trade-off problems is still a challenging subject (Afshar et al. 2009). In situation of contractors competing to finish a given project with the least cost and shortest duration, acquiring the ability to improve the project quality properties seems essential for project managers.

In an attempt to reduce processing time and improve the quality of solutions, particularly to avoid being trapped in local optima, Meta-Heuristic Algorithms (MHAs) have been introduced during last decade (Elbeltagi 2005). MHAs are stochastic search methods that mimic natural biological evolution and social behavior of species. One of the most important criteria of MHAs is their capability of speed convergence to obtain a global optimized solution. Magnetic Charged System Search

Algorithm (MCSS) as a meta-heuristic algorithm proposed by Kaveh et al. (2012) has been applied in this study to improve the convergence ratio of processing and the quality of optimal Pareto Front solutions.

Physics Background
Electrical laws

If we consider charged particles (CPs) as insulating solid spheres of radius a which have a uniform volume charge density and carries a total charge of magnitude q_i, based on physics, the space surrounding these CPs has electric field. The magnitude of the electric field within and outside of the CPs is determined by different expressions (Halliday et al. 2008). Each CP can exert electrical force on any other CP in the space. This force can be determined using Coulomb's law as,According to this law, the magnitude of the electrical force that is exerted on a charge, q_j at position $\mathbf{r}_j$ due the electric field produced by charge q_i at position $\mathbf{r}_i$ can be expressed as,

$$\mathbf{F}_{\mathbf{E},ij} = \begin{cases} k_e \dfrac{q_i q_j}{r_{ij}^2} \dfrac{\mathbf{r}_i - \mathbf{r}_j}{\|\mathbf{r}_i - \mathbf{r}_j\|} & if \quad r_{ij} \le a \\[3mm] k_e \dfrac{q_i q_j}{a^3} r_{ij} \dfrac{\mathbf{r}_i - \mathbf{r}_j}{\|\mathbf{r}_i - \mathbf{r}_j\|} & if \quad r_{ij} > a \end{cases} \tag{1}$$

Where $\mathbf{F}_{\mathbf{E},ij}$ is the electric force that charge q_i exerts on charge q_j and k_e is a constant as the Coulomb constant, and r_{ij} is the separation distance of the two charges and $\mathbf{r}_i$ and $\mathbf{r}_j$ are the positions of the charge q_i and q_j, respectively. In a group of charged particles, the net electrical force that is exerted on charge q_j can be determined by applying the principle of superposition

$$\mathbf{F}_{\mathbf{E},j} = \sum_{i=1, i \ne j}^{n} \mathbf{F}_{\mathbf{E},ij} \tag{2}$$

Where $\mathbf{F}_{\mathbf{E},j}$ is the net electric force that is imposed on the jth CP by other CPs. n is the total number of the charged particles.

Magnetic Laws
Magnetic Fields

The area surrounding a moving charged particle (CP) contains a property that is called magnetic field. Electric currents contain moving charges that induces magnetic fields. For a straight with radius of R which carries an electric current I that is uniformly distributed through the cross section of the wire, using Biot-Savart and Ampère's law, the magnitude of the magnetic field in each point outside and inside of the wire can be obtained as (Halliday et al. 2008),

$$B = \frac{\mu_0}{2\pi} \frac{I}{r} \quad when:\ r \ge R \tag{3}$$

$$B = \left(\frac{\mu_0}{2\pi} \frac{I}{R^2} \right) \times r \quad when: \ r < R$$

Where B is the magnetic field at a point with distance of r from the center line of the wire. If there were a group of wires in the space, the net magnetic field that is induced in a point such as P can be determined as by using superposition principle,

$$B_P = \sum_{i=1}^{n} B_{ip} \tag{4}$$

Where B_p is the total magnetic field at point P, n is the number of wires in the space and B_{ip} is the magnetic field created by the ith wire at point P.

Magnetic Force

When a moving charged particle places in a magnetic field, a magnetic field is imposed on it that can be expressed as,

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B} \tag{5}$$

Where $\mathbf{B}$ is the magnetic field exerted on the moving charged particle and $\mathbf{v}$ is the velocity of the moving charged particle and q is the charge of the particle.

Newtonian Mechanics Laws

Motion of a particle is determined if the particle's position is determined at all times. As a particle moves from an initial position $\mathbf{r}_{old}$ to a new position $\mathbf{r}_{new}$ its displacement, velocity and acceleration can be determined by the following equations,

$$\Delta \mathbf{r} = \mathbf{r}_{new} - \mathbf{r}_{old}$$
$$\mathbf{v} = \frac{\mathbf{r}_{new} - \mathbf{r}_{old}}{t_{new} - t_{old}} = \frac{\mathbf{r}_{new} - \mathbf{r}_{old}}{\Delta t} \tag{6}$$
$$\mathbf{a} = \frac{\mathbf{v}_{new} - \mathbf{v}_{old}}{\Delta t}$$

Considering a constant acceleration that can be determined using second Newton's law, the displacement of the particle can be obtained by,

$$\mathbf{r}_{new} = \frac{1}{2} \frac{\mathbf{F}}{m} \cdot \Delta t^2 + \mathbf{v}_{old} \cdot \Delta t + \mathbf{r}_{old} \tag{7}$$

Magnetic Charged System Search Algorithm

Magnetic Charged System Search proposed by Kaveh et al. (2012) is a multi-agent approach. This algorithm tries to simulate the behavior of a group of charged particles that are moving through a space. In the formulation of the algorithm, some governing physics laws are utilized to properly model the group of charged particles.

Electric Forces

In MCSS algorithm, each CP has an electric charge that is determined based on their objective function fitness as,

$$q_i = \frac{fit(i) - fitworst}{fitbest - fitworst} \tag{8}$$

Where $fit(i)$ is the objective function value of the ith CP, $fitbest$ and $fitworst$ are the so far best and worst fitness among all of the CPs, respectively. So, The electric net force $\mathbf{F}_{E,j}$, acting on the jth CP can be calculated by,

$$\mathbf{F}_{E,j} = q_j \sum_{i,i \neq j} \left(\frac{q_i}{a^3} r_{ij} \cdot w_1 + \frac{q_i}{r_{ij}^2} \cdot w_2 \right) \cdot p_{ji} \cdot (\mathbf{X}_i - \mathbf{X}_j), \begin{cases} w_1 = 1, w_2 = 0 \Leftrightarrow r_{ij} < a \\ w_1 = 0, w_2 = 1 \Leftrightarrow r_{ij} \geq a \\ j = 1,2,...,N \end{cases} \tag{9}$$

Where p_{ji} is the probability of the attraction of the jth CP by the ith CP and r_{ij} their sepration distance (Kave et al. 2012).

Magnetic Forces

It is assumed that the CPs move through virtual wires in the search space. Therefore, the path of the movement of the CPs include some virtual wires. When the CPs move in the virtual wires their charge changes in each movement, so it can be assumed that the virtual wires contain electric current that is,

$$I_{avg} = \frac{\Delta q}{\Delta t} \xrightarrow{\Delta t = 1} \left(I_{avg} \right)_{ik} = q_i^k - q_i^{k-1} \tag{10}$$

Where q_i^k is the electric charge of the ith CP at the end of the kth movement (iteration) and q_i^{k-1} is the electric charge of ith CP at the beginning of the kth movement. In fact, in this algorithm the electric charge represents the amount of the variation of the fitness of each solution after each iteration. For calculating the magnetic forces, we need the velocity of the CPs. Since both CPs in the virtual wire and in the space are moving, the relative speed is used, and since Δt is unity it can be expressed as,

$$\mathbf{v}_{rel} = \mathbf{X}_i - \mathbf{X}_j \tag{11}$$

Now, the total force (Electric and Magnetic) can be determined,

$$\sum \mathbf{F}_j = \mathbf{F}_{B,j} + \mathbf{F}_{E,j} = q_j \sum_{i,i \neq j} \left(\frac{I_i}{r_{ij}} \cdot p_{ji} + (\frac{q_i}{a^3} r_{ij} \cdot w_1 + \frac{q_i}{r_{ij}^2} \cdot w_2) \cdot p_{ji} \right) \cdot (\mathbf{X}_i - \mathbf{X}_j), \begin{cases} w_1 = 1, w_2 = 0 \Leftrightarrow r_{ij} < R \\ w_1 = 0, w_2 = 1 \Leftrightarrow r_{ij} \geq R \\ j = 1,2,..., N \end{cases} \tag{12}$$

Where R is the radius of the virtual wires set to unity. This total force is the source of the movement of the CPs in the search space that helps them converge to global optimum in early stages of the optimization process.

Movement

The movement of these CPs can be calculated based on the aforementioned Newtonian Mechanics laws,

$$\mathbf{X}_{j,new} = rand_{j1} \cdot k_a \cdot \frac{\mathbf{F}_j}{m_j} \cdot \Delta t^2 + rand_{j2} \cdot k_v \cdot \mathbf{V}_{j,old} \cdot \Delta t + \mathbf{X}_{j,old}, \ \mathbf{V}_{j,new} = \frac{\mathbf{X}_{j,new} - \mathbf{X}_{j,old}}{\Delta t} \qquad (13)$$

Where $rand_{j1}$ and $rand_{j2}$ are two random numbers that are uniformly distributed in the range of (0,1). k_a and k_v are the acceleration and velocity factors, respectively. These factors handle the influences of the previous acceleration and velocity (Kaveh and Talathari 2009). m_j is the mass of the particle that is set to q_j in this paper. Δt is the time intervals that is set to unity.

TCO Formulation

TCO model tries to find the optimal solutions considering two objectives that Cost and Time of the project. So, it is a multi-objective optimization that simultaneously concerns two objectives. The objective of the time of project can be determined as,

$$ES_j = \max_{i \in z_j} \{EF_i\} \ and \ EF_i = ES_i + t_{i,j}.y_{i,j} \qquad (14)$$

$$1 = \sum_j y_{i,j} \qquad (15)$$

$$Z_t = \max_{\forall i} \{EF_i\} \qquad (16)$$

Where ES_i and EF_i are the earliest start time and earliest finish time of activity i, respectively. z_j is the set of predecessors of j activity. So, Equation (14) considers the precedence constraint. $t_{i,j}$ is the time of activity i using the j option for this activity. $y_{i,j}$ is a binary decision variable, and it is equal to unity if activity i are performing by option k, and in other cases it is zero. Equation (15) considers the constraint that only one option should be utilized for an activity. By this formulation, the total time of the project can be calculated by Equation (16).

The other objective of the optimization is the cost of the project that can be divided in two categories that are direct and indirect costs. Direct costs depend on the activities and the option used for them, but the indirect costs significantly depend on the duration of the activities. The total cost of the project can be expressed as,

$$DC = \sum_{\forall i} dc_{i,j} \cdot y_{i,j} \ and \ IC = T \times ic \qquad (17)$$

$$Z_c = DC + IC \qquad (18)$$

Where DC is the direct cost of the project, $dc_{i,j}$ is the direct cost of the activity i with the option k. IC is the indirect cost of the project and ic is the indirect cost rate of the project. Z_c is the total time of the project that is sum of direct and indirect costs.

Modified Adaptive Weight Approach

For dealing with these two objectives simultaneously, we used Modified Adaptive Weight Approach (MAWA) developed by Zheng et al. (2004). According to MAWA, the adaptive weights are formulated based on the following four conditions:

1. For $z_t^{max} \neq z_t^{min}$ and $z_c^{max} \neq z_c^{min}$,

$$v_c = z_c^{min}/(z_c^{max}-z_c^{min}) \ and \ v_t = z_t^{min}/(z_t^{max}-z_t^{min}) \qquad (19)$$

$$v = v_t + v_c \qquad (20)$$

$$w_c = v_c/v \ and \ w_t = v_t/v \qquad (21)$$

2. For $z_t^{max} = z_t^{min}$ and $z_c^{max} = z_c^{min}$,

$$w_t = w_c = 0.5 \qquad (22)$$

3. For $z_t^{max} = z_t^{min}$ and $z_c^{max} \neq z_c^{min}$,

$$w_t = 0.9 \ and \ w_c = 0.1 \qquad (23)$$

4. For $z_t^{max} \neq z_t^{min}$ and $z_c^{max} = z_c^{min}$,

$$w_t = 0.1 \ and \ w_c = 0.9 \qquad (24)$$

Where z_t^{max} and z_c^{max} are the maximal values of the duration and total cost of the project in the current iteration. z_t^{min} and z_c^{min} are the minimal values of the duration and total cost of the project in the current iteration. v_c is the value for criterion of cost, and v_t is the value of the criterion of time, and v is the value of the project. At last, w_t and w_c are the adaptive weight for the time and total cost objectives, respectively. Using presented adaptive weights the fitness function can be formulated as,

$$f(x) = w_t \frac{z_t^{max} - Z_t + \gamma}{z_t^{max} - z_t^{min} + \gamma} + w_c \frac{z_c^{max} - Z_c + \gamma}{z_c^{max} - z_c^{min} + \gamma} \qquad (25)$$

Where γ is a small positive random number between 0 and 1 that is defined to avoid a zero division. Considering the Equation (27) as the objective function, we can utilize MCSS algorithm to solve the problem, and each CP will be a feasible solution of the problem.

Pareto Front
In Time-Cost optimization, it is important to provide a complete Pareto front. A complete Pareto front will give decision makers more freedom in selecting appropriate construction time and cost. Since the objective of this optimization is to minimize both objectives, we utilize the basis of Pareto dominance. If the following condition is met, a decision vector u is said to strictly dominate the other v:

$$D_u < D_v \ and \ C_u \leq C_v \ or \ D_u \leq D_v \ and \ C_u < C_v \qquad (26)$$

Where D_u and D_v are durations of u and v decisions, respectively. Also, C_u and C_v are cost of u and v decisions, respectively. A set of decision vectors that none of its members can be dominated by any others are nondominated set. The Pareto front is also a nondominated set. Using the aforementioned concepts, the points that are in the Pareto front can be determined.

Numerical Example

The numerical example is fitted in the prototyped mode to verify the performance of the proposed model. An example project with seven activities is utilized for this purpose. Figure 1 shows the activity on nodes (AON) diagram of this example. The options available for each activity and their corresponding duration and direct cost are described by Zheng et al (2004). The indirect cost of this project is set to $1,500 per day. The number of CPs is set to 5 and the maximum number of iterations is set to 10. The optimum solution found by MCSS algorithm indicates that by considering the fitness function of Equation (25) the optimum solution with cost of $233,500 and duration of 60 days. This solution is the first solution that is listed in Table 1. Also, the other Pareto optimal solutions found by the MCSS algorithm during the optimization process are shown in Table. 1. For a comparison, the Pareto optimal solutions that are provided by the other methods proposed by Zheng et. al (2004) and Gen and Cheng (2000) are also listed in Table 1. In Table 1, the numbers of the activities determine the option that is selected for each activity. It can be seen that MCSS algorithm can provide better Pareto optimal solutions. This means that MCSS has a more efficient performance comparing to GAs algorithms.
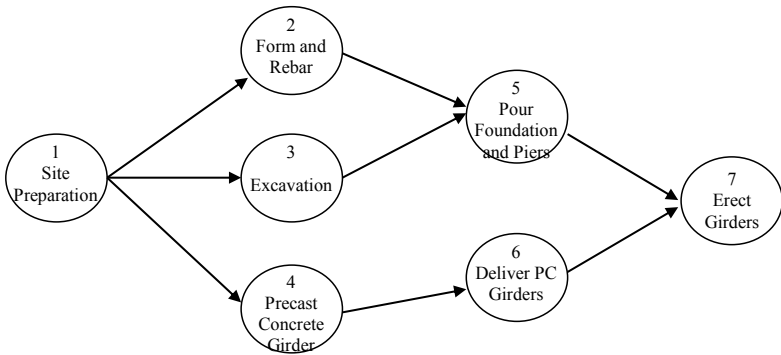
Figure 1. Activity on Node (AON) Project Network

Table 1. Pareto front Solutions of the numerical example

Approaches	Number	Activity							Time	Cost
Gen and Cheng (2000)	1	3	2	1	2	1	1	2	79	256,400
	2	2	4	1	2	1	3	3	83	243,500
	3	3	2	2	2	1	3	3	86	242,400
Zheng et al. (2004)	1	1	1	1	3	3	2	1	66	236,500
Present Work	1	1	1	1	1	1	3	1	60	233,500
	2	1	1	1	3	2	2	1	62	233,000
	3	1	1	1	3	3	3	1	67	224,000

Conclusions

The methodology of the MCSS algorithm as a Meta-Heuristic algorithm is described and applied to TCO problem. MCSS algorithm emphasizes on the exploration ability at the early stages of the optimization process and the exploitation ability in the last stages. This causes a perfect search and convergence ability. Also, the existence of the magnetic forces enables the search agent to save their experience in their last movement, and transfer it to other agents. The result of the numerical example adapted from Zheng et al. (2004) verifies the efficiency of the MCSS algorithm for TCO problem. MCSS has been outperformed other approaches that used variations of GA algorithm for the optimization process.

References

Afshar, A., Kasaeian Ziaraty, A., Kaveh, A., Sharifi, F. (2009). "Nondominated Archiving Multicolony Ant Algorithm in Time-Cost Trade-Off Optimization." *J. Constr. Eng. Mgmt.,* 135(7), 668-674.

Gen, M., and Cheng, R. (2000). "Genetic algorithms & engineering optimization." *Wiley-Interscience*, New York.

Halliday, D., Resnick, R., and Walker, J. (2008). "Fundamentals of physics." 8th edn. *Wiley*, New York.

Kaveh, A., and Talathari, S. (2010). "A novel heuristic optimization method: charged system search." *J. Acta Mech*., 213(3-4), 267-289.

Kaveh, A., Motie Share, Mohammad A., and Moslehi, M. (2012). "Magnetic charged system search: a new meta-heuristic algorithm for optimization." *J. Acta Mech*., 224(1), 85-107.

Zahraie, B., and Tavakolan., M. (2009). "Stochastic Time-Cost-Resource Utilization Optimization Using Non-dominated Sorting Genetic Algorithm and Discrete Fuzzy Sets." *J. Constr. Eng. Mgmt*., 135 (11), 1162-1173.

Zheng, D. X. M., Ng, S. T., and Kumaraswamy, M. (2004). "Applying a genetic algorithm-based multiobjective approach for time–cost optimization." *J. Constr. Eng. Manage.*, 130(2), 168–176.

Construction Scheduling and Resource Allocation based on Actual State Data

T. Horenburg[1] and W. A. Günthner[1]

[1]Institute for Materials Handling, Material Flow, Logistics, Technische Universität München

ABSTRACT

Construction projects are subject to unexpected factors such as weather, soil conditions etc., which cause frequent changes in baseline schedules. Yet existing precedence diagrams (PERT, CPM) can not provide reliable results to handle the complexity of an ideal resource allocation (Resource Constrained Project Scheduling Problem) – especially when considering actual state data: Initial assumptions become obsolete and disturbances in process durations or sequence cause iterative revisions of the schedule. Therefore a method is introduced to optimize construction schedules continuously based on the current progress considering all relevant resource capacities. Actual state data is progressed and variations in level of detail between construction works and the virtual model are leveled. A Multi-Agent Framework generates valid schedules based on heuristic priority rules to minimize project makespan and additionally determines floats and total resource costs. Thus various scenarios can be analyzed to meet the specific requirements of the project and to assess the consequences of schedule changes more effectively.

INTRODUCTION AND RELATED WORK

Low productivity rates and frequent changes in baseline schedules often characterize construction projects. Respective consequences in terms of resource allocation for future activities are hardly considered. Although existing precedence diagrams such as PERT or CPM provide a certain basis for computational project management, complexity and resulting coordination effort affect construction works heavily. Usually processes require resources such as machinery, material and space, which are typically scarce and therefore provide additional constraints to the common predecessor-successor relationships. In research the ideal allocation of limited resources to multiple processes (NP-hard) is known as Resource Constrained Project Scheduling Problem (RCPSP) (Brucker et al., 1999), which aims for minimum project makespan.

Valid schedules however are subject to unexpected factors, e.g. weather, soil conditions, bottlenecks in supply etc. Hence initial assumptions become obsolete and major disturbances in process durations or sequence cause iterative revisions of the optimized baseline schedule. The latter are mainly depending on the skills and

rationality of the construction manager, who has to reschedule in order to avoid cost intensive idle time of machines and human resources.

Branch-and-bound methods provide an exact solution of the RCPSP (Brucker et al, 1998), though the computational effort is hardly acceptable for the number of processes involved in common construction projects. For this reason different heuristics and meta heuristics were developed to meet the specific requirements of the RCPSP. These procedures are usually based on the determination of processes executable – prior tasks are completed and required resources available – and a respective resource allocation at each event caused by the completion of a task. The priority of individual processes and the respective sequence is determined at random (constraint-based Monte-Carlo-Simulation) (König et al, 2007), based on soft constraints considering execution strategies or by different meta heuristics such as particle swarm optimization (Lu et al., 2008) or genetic algorithms (Toklu, 2002). Further approaches apply priority rules to reduce the number of iterations to provide acceptable solutions in a basic software environment (Kolisch, 1996) or in an agent-based decision system (Knotts et al., 2000). The RCPSP can be specified according to Brucker's et al notation and adapted to the approach introduced in this paper:

- set of activities $V \in \{0,1\ldots n+1\}$
- set of renewable resources $R \in \{1,2\ldots m\}$
- set of non-renewable resources $M \in \{1,2\ldots d\}$
- activity/job $j \in [V]$ with duration $p_j > 0$
- renewable resources $k \in [R]$ with amount of available resources R_k
- non-renewable resources/material $h \in [M]$ with amount M_h
- renewable resources r_{jk} required by activity j from resource k
- non-renewable resources h_{jk} required by activity j from resource h
- set of predecessors P_j of activity j
- set of successors S_j of activity j
- capital c_j of activity j (Brucker et al., 1999)

An additional dummy activity is defined both for the start and the end of the project requiring no resources and time ($p_{0 / n+1} = 0$; $r_{0k / n+1k} = 0$).

Typically construction schedules and respective resource allocations are generated manually. Yet in research pattern-based approaches provide project plans based on models such as Building Information Models (BIM) (Benevolenskiy et al., 2012; Wu et al., 2010; Olbrich et al., 2012). Therefore predecessor-successor relationships are derived from hierarchical templates. Specific parameters for individual instances are set to define required resources and further constraints. The generated schedules however can not handle limited resources, thus a realistic project plan for construction is not defined.

      Research activities also focus on data acquisition to determine the current construction progress. Therefore various sources such as laser scanning, photographic pictures, sensors or machine data are utilized for an automated acquisition. The determination of the actual state or progress of specific processes is very challenging, but should not be further evaluated within this publication.

Few approaches consider the improvement of the process sequence once construction works have started. AbouRizk introduced a generic platform (COSYE) that facilitates

collaborational development (re-)using various data sources for construction simulation (AbouRizk et al., 2011). The general application of actual state data was successfully analyzed by Ailland and Bargstädt (2010). Marx and König (2011) used a similar method to model the current spatial conditions based on three-dimensional geometric constraints.

Even though the capability of the gained progress data is very high, the according project-relevant information is rarely used for further optimization of project schedules. Hence implications from schedule changes or delays are hard to estimate and remaining construction processes may interfere by exceeding resource capacities in future sections resulting in further unexpected delays. Therefore the approach introduced in this paper optimizes the schedule continuously based on the current construction progress considering all relevant resource capacities.

MULTI-AGENT FRAMEWORK

The applied concept is based on prior works in Multi-Agent-Systems (MAS), yet both processes and resources are modeled as autonomous agents. Valid schedules are generated by authorizing agents to negotiate as soon as all existing preconditions are satisfied (figure 1: P1.2, P2.2). Process agents then request proposals on a central blackboard (negotiation protocol), where all agents may negotiate on resource allocation whilst considering existing constraints. Admitted resources (A, B) propose offers according to their specific qualification (abilities, location, costs, etc.) for the jobs available. Depending on defined target values and bidding algorithms resources are allocated to prioritized processes (P2.2) ensuring ideal construction sequences.
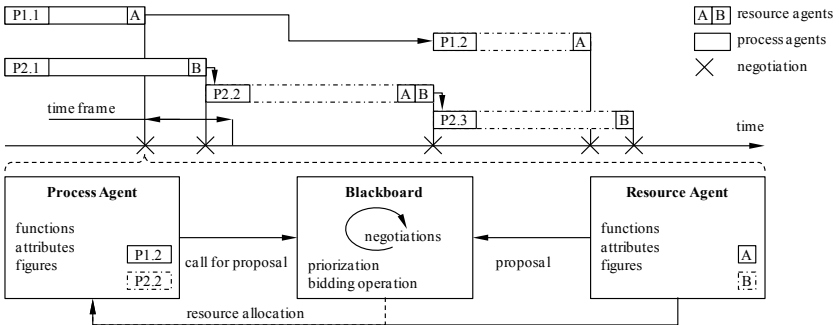
Figure 1. Negotiations in a Multi-Agent Framework.

The implementation of the introduced approach is based on discrete events in a simulation environment (DES). Therefore critical activities may be analyzed on a higher level of detail resulting in exact process durations of major validity for any specific process agent. Agents as such can be defined as self-contained program aware of its own decisions that operates according to its perception of the environment pursuing one or multiple objectives. Smart Agents comply with at least two of the three criteria listed in table 1.

Table 1. Smart Agents (Ren and Anumba, 2004).

Criterion	Description
Autonomy	Agents operate without human guidance chasing specific objectives. An essential element is pro-activeness, i.e. the initial ability to operate itself rather than merely reacting.
Cooperation	This characteristic is compulsory for MAS. Interaction between individual agents and possibly humans is necessary. For this purpose a communication language is set up.
Learning	From reactions and/or interactions with their environment intelligent entities learn, so they can act ideally in similar situations

Each Process agent represents an activity of the respective schedule containing all necessary information such as capital c_j, duration p_j or start and completion dates. Constraints are provided as set of predecessors P_j and successors S_j as well as for (non-)renewable resources r_{jk} (h_{jk}). Control mechanisms for communication and interaction are responsible for any information transfer. Table 2 shows all states of process agents and the corresponding operations. The specified states vary according to the conditions of the predecessors and the availability of the required resources.

Table 2. States of the implemented process agents (Horenburg et al, 2012).

state	description	operations
blocked	preconditions not met	wait for predecessors
admitted	preconditions met for negotiations	registration at blackboard & call for proposals
accepted	resources allocated	wait for resources & predecessors' completion
active	resources allocated & working	finish activity in duration p_j, update successors
complete	activity successfully finished	deallocation of resources

Both processes and resources are implemented as intelligent entities with parameters, functions and objectives. Thus resource agents are able to participate in the decision-making by offering their labor according to specific capability functions. Generally there are two types: renewable resources such as machinery, workers or space are available for every period whilst non-renewable resources are only available with a defined amount of units (e.g. material). Each entity is an instance of either renewable or non-renewable resource agents differing in attributes and activities performable. All resource agents imply the three states free, active and reserved.

For resource allocation admitted process agents call for proposals on a central market place requesting renewable resources r_{jk} and non-renewable resources h_{jk} to complete their activities as early as possible. The order of process agents negotiating on resource allocation is defined by priority rules, which were implemented in the simulation environment (e.g. Longest Path Following, Minimum Slack, Worst Case

Slack). Furthermore a time frame was introduced (see figure 1) to allow the reservation of resources by high risk process with predecessing activities not completely finished (Horenburg et al, 2012). Resource agents fulfilling the requested criteria offer their labor depending on the qualification for the quoted activity. Thus proposals are dynamically regulated by the state of the individual resource as well as its attributes and the influences of the environment. Once allocated resources stay with the corresponding process agent until completion of the activity. An implemented observer evaluates workload and costs for individual resources or a group of resources as well as overall project costs.

The general problem of makespan minimization is split into decisions of less complexity. The decentralized agent objectives should lead to a general improvement of the system. The current approach was evaluated with standardized schedules generated by the project generator ProGen (Kolisch and Sprecher, 1996). For 480 instances of n = 120 activities mean deviation from optimum was about 8% (priority rule: Longest Path Following). Furthermore the approach provided better solutions than Monte-Carlo simulation whilst the required computing time being around 500 times lower (Horenburg et al., 2012).

Having shown the efficiency of the introduced MAS for generating feasible and good solutions of baseline schedules, the method is applied to update schedules during construction works.

ACTUAL STATE DATA INTEGRATION

The infrastructure and the respective robustness of the developed MAS allows the import of current progress data to determine optimized process sequences for remaining construction processes (see figure 2). Actual state data from various sources can be computed in a specified format. The progress of the individual activities is determined once new data is available.
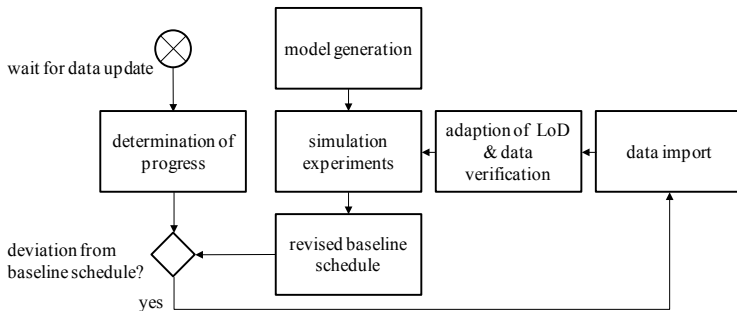
Figure 2. Concept for project scheduling based on actual state data.

The MAS distinguishes between finished processes, activities in progress and processes that have not started yet. For the former start and finishing times are known, for the latter merely predicted durations from the baseline schedule (i.e. estimations based on experience). Activities in progress however are represented by active process agents and corresponding resources. Depending on the level of detail (LoD) of the utilized input data new activities might be generated (high LoD) or

combined to summary tasks (low LoD). The import function is able to handle differences in the LoD of progress information on the construction site and the corresponding virtual model. For scarce information the progress of multiple processes must be determined from insufficient data. Especially activities that are not on the critical path of the partial project plan (see figure 3 left) are subject to respective uncertainties. Yet the latter can be compensated by validity checks in terms of resource allocation and further simulation experiments.

$$LoD_{real} < LoD_{sim} \qquad LoD_{real} = LoD_{sim} \qquad LoD_{real} > LoD_{sim}$$
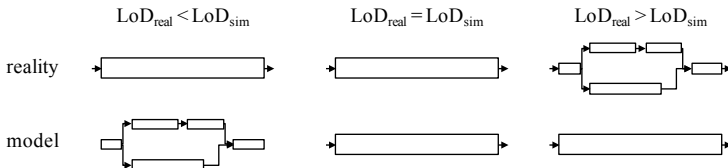
Figure 3. Levels of Details for actual state data and simulation.

For construction activities that can be described by constant progress functions, the progress in percent equals the temporal progress. For linear functions (e.g. earthworks involving increasing transport times) the temporal progress can also be determined – provided that the duration of the first as well as the last work step is known. Yet for more complex tasks involving non-linear progress specific functions have to be implemented. A precise data verification is critical to avoid and/or detect mistakes in data aggregation. Due to the problem specification predecessing activities have to be finished to start succeeding activities. Hence processes that are either marked as finished or active are analyzed in terms of consistency in their predecessor relationships. Furthermore the total number of active resources is identified and compared to the original data from the baseline schedule.

Based on these information start times of running activities are set and remaining durations are updated. Respective successors are eligible for agent negotiations and the remaining construction period can be revised with various simulation experiments.

RESULTS

Depending on the general conditions of the construction project different scenarios can be analyzed and compared to the baseline schedule. Further resources can be added or eliminated to complete all activities in time or to reduce resource costs respectively. Impacts both on makespan and costs are computed for the remaining constructions activities and the utilization of all resource agents is determined. Additionally free and total floats are identified for every process by backward scheduling considering resource constraints.

Figure 4 shows a sample project of 30 activities with an initial baseline schedule and a revised schedule after 22 days. Original makespan was 46 days resulting in total cost of 48.144 monetary units (mu). Due to variations of processes during construction a new sequence was created to minimize the remaining project makespan without any additional changes regarding number of resources or individual activity duration.

Initial Plan			Rescheduling after 22 days		
Duration [d]	46		Duration [d]		57
Total costs [mu]	48.144		Total costs [mu]	20.844 + 40.184	= 61.028
fixed	41.768		fixed	18.484 + 36.272	= 54.756
variable	6.376		variable	2.360 + 3.912	= 6.272
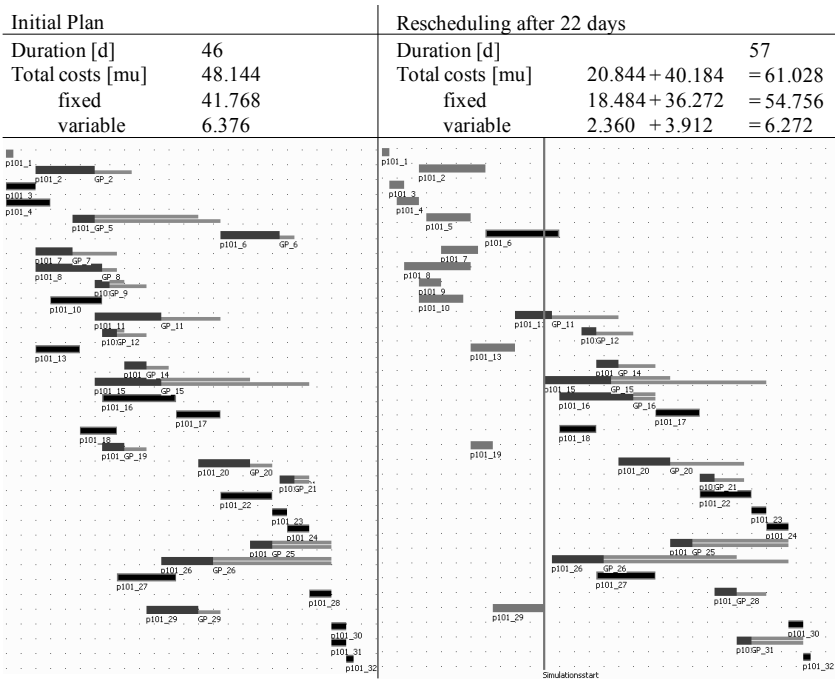
Figure 4. Schedule update – example.

Since actual durations of finished and active processes differ from initial assumptions total variable costs are lower than expected. Floats alter and critical activities change respectively. To reduce the total makespan of the project further resources could be added or durations of specific (critical) processes could be decreased by changing the construction method and/or resource requirements.

As a result of MAS negotiations an optimized construction schedule is generated considering current progress data.

CONCLUSIONS

The concept introduced in this paper improves the scheduling of construction projects by ensuring an intelligent allocation of resources based on MAS. The possibility of updating activities and schedules during construction enables an optimization of the remaining activity sequence based on current progress information. Subject of optimization can both be minimal project makespan as well as minimal resource deployment to complete all construction tasks in time. Furthermore existing floats as well as total costs of resources are calculated. Thus various scenarios can be analyzed to meet the specific requirements of the project and to assess the consequences of schedule changes more effectively.

REFERENCES

AbouRizk, S., Halpin, D., Mohamed, Y. and Hermann, U. (2011). Research in Modeling and Simulation for Improving Construction Engineering Operations. *In: Journal of Construction Engineering and Management*, 137(10), 843-852.

Ailland, K. and Bargstädt, H. J. (2010). Construction Process Simulation Based on Signifcant Day-to-Day Data. *In: Proc. Of the 13th International Conference on Computing in Civil and Building Engineering*, Nottingham.

Benevolenskiy, A., Roos, K., Katranuschkov, P. and Scherer, R.J. (2012). Construction processes configuration using process patterns. *In: Advanced Engineering Informatics*, 26(4), 727-736.

Brucker, P., Knust, S., Schoo, A. and Thiele, O. (1998), A branch and bound algorithm for the resource-constrained project scheduling problem. *In: European Journal of Operational Research,* 107(2), 272–288.

Brucker, P., Drexl, A., Möhring, R., Neumann, K. and Pesch, E. (1999). Resource-constrained project scheduling: Notation, classification, models, and methods. *In: European Journal of Operational Research*, 112, 3-41.

Horenburg, T., Wimmer, J. and Günthner W. A. (2012). Resource Allocation in Construction Scheduling based on Multi-Agent Negotiation, *In: Proc. of the 14th Int. Conference on Computing in Civil & Building Engineering,* Moscow.

Kolisch, R. (1996). Efficient priority rules for the resource-constrained project scheduling problem. *In: Journal of Operations Management*, 14, 179-192.

Kolisch, R. and Sprecher, A. (1996). PSPLIB – A project scheduling library. *In: European Journal of Operational Research*, 205-216.

König, M., Beißert, U., Steinhauer, D. and Bargstädt, H.-J. (2007). Constraint-Based Simulation of Outfitting Processes in Shipbuilding and Civil Engineering. *In: 6th EUROSIM Congress on Modeling and Simulation,* Ljubljana.

Knotts, G., Dror, M. and Hartman, B. C. (2000). Agent-based project scheduling. *In: IIE Transactions*, 32, 387-401.

Lu, M., Lam, H.-C., Dai, F. (2008). Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization. *In: Automation in Construction*, 17(6), 670-681.

Marx, A. and König, M. (2011). Preparation of Constraints for Construction Simulation. *In: Proc. of the 2011 ASCE International Workshop on Computing in Civil Engineering*, Miami.

Olbrich, L., Enge. F. and Huhnt, W. (2012). Hierarchical templates for the design of planning and construction processes. *In: Proc. of the 14th International Conference on Computing in Civil and Building Engineering,* Moscow.

Ren, Z. and Anumba, C.J. (2004). Multi-agent systems in construction-state of the art and prospects. *In: Automation in Construction*, 13(3), 421–434.

Toklu, Y.C. (2002). Application of genetic algorithms to construction scheduling with or without resource constraints. *In: Canadian J. of Civil Eng.*, 29(3), 421–429.

Wu, I.-C., Borrmann, A., Beißert, U., König, M. and Rank, E. (2010). Bridge construction schedule generation with pattern-based construction methods and constraint-based simulation. *In: Advanced Eng. Informatics*, 24(4), 379-388.

Tracking Secondary and Temporary Concrete Construction Objects Using 3D Imaging Technologies

Y. Turkan[1], F. Bosché[2], Carl T. Haas[3], Ralph Haas[4]

[1]Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, Iowa, 50011-3232; PH (515) 294-7539; FAX (515) 294-3845; email: yturkan@iastate.edu
[2]School of the Built Environment, Heriot-Watt University, Edinburgh, Scotland; PH 44 (0) 131 451 4659; email: f.n.bosche@hw.ac.uk
[3]Department of Civil Engineering, University of Waterloo, Ontario, N2L 3G1, Canada; PH (519) 888-4567 ext. 35492; FAX (519) 888-4300; email: chaas@uwaterloo.ca
[4]Department of Civil Engineering, University of Waterloo, Ontario, N2L 3G1, Canada; PH (519) 888-4567 ext. 32176; FAX (519) 888-4300; email: haas@uwaterloo.ca

ABSTRACT

Recent research efforts to improve construction progress tracking has focused on employing emerging technologies such as three dimensional (3D) imaging, including digital photogrammetry and 3D Terrestrial Laser Scanning (TLS). Previous research has shown that "Scan-vs-BIM" object recognition systems, which fuse 3D TLS and 4D project BIM, provide valuable information for tracking structural works. However, until now these systems have focused on tracking progress for permanent structures only; none of them has considered progress of secondary or temporary structures. In the context of structural concrete work, temporary structures include formwork, scaffolding and shoring, while secondary components include rebar. The value of tracking temporary and secondary elements is that it would add veracity and detail to the progress tracking process, and consequently to billing. This paper presents two different techniques for detecting concrete construction secondary and temporary objects in TLS point clouds, one of which is based on a Scan-vs-BIM object recognition system. Both techniques are tested using real-life data collected from a reinforced concrete building construction site. The preliminary experimental results show that it is feasible to detect and track secondary and temporary objects in 3D TLS point clouds with high accuracy. This will help to improve progress estimation and tracking.

Keywords: **Construction progress tracking, Laser scanning, BIM, Object recognition, Temporary objects, Secondary objects**

INTRODUCTION

Accurate and effective progress tracking is a must for successful management of construction projects as it allows corrective decisions to be made in a timely fashion. Traditional construction progress tracking methods involves manual data collection and data extraction from different construction documents which is tedious and time consuming. Recent research efforts to improve progress tracking are mainly focused on employing technologies such as three dimensional (3D) imaging including digital photogrammetry (Golparvar-Fard et al. 2009; Wu et al. 2010) and 3D laser scanning (Cheok et al., 2000; Bosché and Haas 2008; Turkan et al. 2012). However, none of these systems report progress of secondary or temporary structures, i.e. their focus is mainly on tracking permanent structure's progress. Nonetheless, secondary and temporary structures' progress would add veracity and detail to the progress tracking process. Temporary construction objects such as formwork, scaffolding, and shoring are the largest cost components of a concrete building's structural frame (Hurst, 1983; Peurifoy and Oberlender, 2011). Together with the secondary objects such as rebar, total cost of temporary and secondary objects constitute a significant portion of a concrete building's structural frame's cost (Jarkas and Horner, 2011). Therefore, it is important to track these elements to increase the accuracy of progress tracking and also better support billing.

This paper describes two techniques for recognizing concrete construction secondary and temporary objects in TLS point clouds, one of which based on an object recognition system developed by Bosché and Haas (2008) that uses a "Scan-vs-BIM" framework. Both techniques are tested using real life data that is obtained from a reinforced concrete building construction.

LITERATURE REVIEW

The literature review revealed that at present there are only a limited number of studies on using 3D imaging technologies for automated detection and tracking of secondary and temporary construction objects. Two relevant research works have been identified by the authors. One is by Lee et al. (2010) who developed an image-based technique for calculating the quantity of formwork installed from construction site images. Although high recognition values were reported in this work (as much as 90%), there are issues with sunlight, shadow, obstructions etc. The second relevant research is the work by Ishida et al. (2012) who developed a system to inspect the quality of a reinforced concrete structure using 3D point clouds obtained with TLS. Essentially, they used a shape recognition technique to detect steel rebars in reinforced concrete structures. Their system successfully identified the rebars in the 3D point clouds, and was able to count the number of column ties and vertical ties.

The Automated Object Recognition System: One of the secondary and temporary element recognition techniques described herein is built upon the object recognition algorithm proposed by (Bosché and Haas, 2008) to recognize designed 3D model objects in TLS point clouds. This "Scan-vs-BIM" system and its experimentally validated performance are detailed in (Bosché and Haas, 2008; Bosché, 2010). A brief review is given below:

Registration of TLS Point Clouds with Building 3D Model: An initial coarse registration is performed, for example by manually matching n pairs of points selected in the 3D model and in the scan using commercially available cloud processing software. A robust Iterative Closest Point (ICP) based algorithm using the point-to-plane framework (Chen and Medioni, 1991; Rusinkiewicz and Levoy, 2001) is then employed to perform the fine registration of the TLS point clouds with the building 3D model. Following that for each scanned data point, a matching model point is calculated as the closest of the orthogonal projections of the data point on the objects' triangulated facets. For ensuring robustness of the matching and consequently registration with respect to outliers, point pairs are rejected when:

(1) The Euclidean distance between two matched points is larger than a threshold τ_D. We typically use τ_D =20mm for structural elements.

(2) The angle between the normal vectors to two matched points is larger than a threshold τ_A. We typically use $\tau_A = 45^o$.

The ICP iterative process is stopped when $\Delta MSE < 0.05 mm^2$, where ΔMSE is the improvement in Mean Square Error (MSE) between the current iteration and the previous one.

Recognition of 3D Model Objects in TLS point clouds: At the end of the registration process, the project 3D model and TLS point clouds are optimally registered. Because it is known to which object points were matched at the last iteration, each model object can be assigned a corresponding as-built point cloud. The analysis of the as-built point cloud can then lead to the recognition of the object itself using a surface-based recognition metric (Bosché, 2010). The percentage of recognition $\%_{recognized}$ is calculated as:

$$\%_{recognized} = \frac{S_{recognized}}{S_{recognizable}} = \frac{S_{recognized}}{S_{planned} - S_{occluded}}$$

where $S_{planned}$, $S_{occluded}$ and $S_{recognized}$ are the planned, occluded and recognized surfaces. $\%_{recognized}$ and $S_{recognized}$ are used to infer the recognition of each object. We typically use the rule:

If ($\%_{recognized}$ >50% or $S_{recognized}$ >1000cm^2),
     then the object is considered recognized.

TECHNIQUES FOR CONCRETE CONSTRUCTION TEMPORARY AND SECONDARY OBJECTS RECOGNITION AND TRACKING

We describe two techniques here for recognizing concrete construction secondary and temporary objects in 3D TLS point clouds.

Technique 1: The first technique simply changes the default point matching metrics in the approach of Bosché and Haas (2008). It is hypothesized that formworks and rebar could be detected by searching for matching points that are slightly further away than the default τ_D value. Furthermore, the τ_A threshold is discarded to enable

the recognition of formworks used for forming the faces of walls or columns hidden from the scanner's location. The recognition of column formwork (and rebar) and columns themselves should then be differentiable by analyzing the variation of $\%_{recognized}$ over a range of values of τ_D, e.g. between 10mm to 60mm (Figure 1). A completed column should present high recognition levels for most values of τ_D (10-20mm). Formworks typically have thicknesses of 30mm and more, while the concrete cover in an RCC column and beams is typically 40mm (1.5in) (Nunnally, 2004). As a result, formworks and rebar should lead to low recognition levels for low τ_D values, and higher recognition levels only for large τ_D values. However, because rebar has a visible surface that is smaller than the finished object, recognition levels for larger τ_D values should be lower than for formworks. Technique 1 can be applied to detect formwork and rebar, but cannot be used for shoring.

(a)   **(b)**   **(c)**
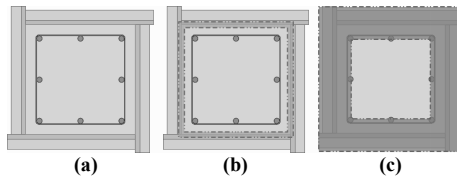
Figure 1: Formwork and rebar recognition using Technique 1: (a) top view of a concrete column showing the reinforcement and formwork; (b) Recognition of finished concrete using a small value of τ_D e.g. 20 mm (green volume) (c) Recognition of rebar and formwork using a large value of τ_D, e.g. 50mm, (red volume)

Technique 2: An "open space" volume is an empty 3D space volume in the 3D model. For example, the open space volume for floor slab shoring can be defined as the cubic space surrounded by four columns. Given an open space volume, the total number n of TLS data points contained within that volume can be calculated, and the number of points per cubic meter η inferred. Then, if $\eta > \eta_{min}$, shoring can be considered detected. This technique can be applied to detect shoring. Its advantage for shoring is simplicity and effectiveness (as shown later). However, its disadvantage is the necessity to run a routine on the 3D model which can interpret it automatically to create the volumes to be analyzed. This should be generally straight forward, but complex design geometries could present challenges.

EXPERIMENTS

In order to evaluate the performance of the proposed secondary and temporary construction object recognition techniques, experiments were conducted using a 3D model obtained for the Engineering V Building site at the University of Waterloo, Canada and eleven different TLS point clouds acquired on eleven different days over a seven month period (Figure 2). The 3D laser scans were acquired using a Trimble® GX 3D laser scanner that uses time-of-flight technology.

(a)   (b)   (c)
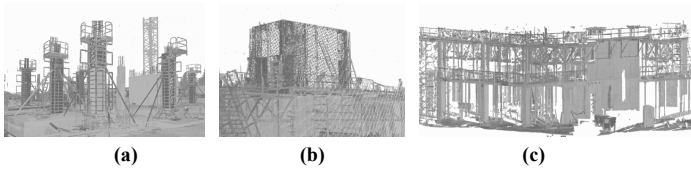
Figure 2: Temporary and Secondary objects at E5 Building site: (a) column formworks – July 25, 2008; (b) Rebar – September 19, 2008; (c) Slab formworks and shoring – October 9, 2008.

Recognition of Completed Objects (Step 1): In order to recognize secondary and temporary structures, the algorithm of Bosché and Haas (2008) is first run using its default settings, i.e. τ_D =20mm, $\%_{recognized}$ =50% and $S_{recognized}$ =1,000cm^2. This step leads to the recognition of all completed elements, and subsequently the removal from the TLS point clouds of all the points matched to the recognized elements.

Recognition of Formwork and Rebar (Step 2): As a second step, Bosché and Haas' (2008) algorithm is run with values of τ_D ranging from 10mm to 60mm in order to identify whether rebar and formwork could be distinctly differentiated from completed objects. Five of the TLS point clouds were used here which collectively contain data from 111 column, floor and wall objects in different construction states, namely: "built" (i.e. completed), "formwork" and "rebar". The experiment results are summarized in Figures 3 for columns and walls respectively. Overall, it appears, as expected, that recognition levels for objects in "formwork" state show very low $\%_{recognized}$ for values of τ_D <30mm, but then clearly increase from 30mm onwards. Objects in "rebar" state show similar patterns as those in "formwork" state but do not reach as high recognition levels for larger values of τ_D, which was also expected. Therefore, it can be concluded that the analysis of the variation of $\%_{recognized}$ over the suggested range of values for τ_D can be used to distinguish the different construction states of concrete structures.

Figure 4 shows an example of the results obtained for a scan after the first two steps are successively applied.

(a)   (b)

Figure 3: Recognition performance for different values of τ_D and at different construction stages. The curves show the average and standard variation of $\%_{recognized}$: (a) columns (b) walls

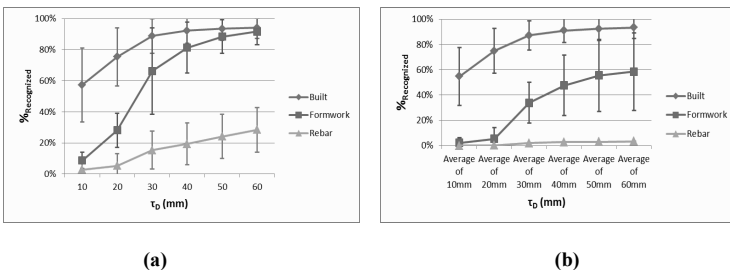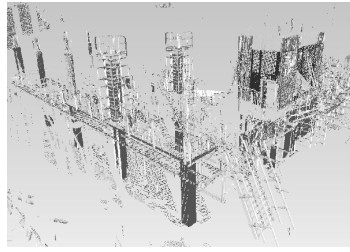**Figure 4 Illustration of the recognition of concrete structure elements in state "built" (blue), "formwork" (red) and "rebar"(green).**

Recognition of Shoring (Step 3): In order to recognize shoring in TLS point clouds, a simple technique based on the analysis of points contained within open spaces defined by design objects, is tested. Trimble® Realworks® manual segmentation tool was used to select sets of points from the point cloud by defining boundaries using its polygonal framing function; the tool reports the number of points in the defined volume. Separately, the corresponding volume of the defined volume is calculated using the structural BIM model and commercial BIM software (Autodesk® Revit® was used here). While our experiment performs this task manually, full automation is feasible. 50 open spaces that have shoring and another 50 open spaces that do not have shoring were selected from the TLS point clouds of the Engineering V Building. Figure 5 shows that the number of points per cubic meter varies between 20 and 40 if there is no shoring, and between 60 and 100 if there is shoring in the selected open space volume (these results were obtained with a relatively constant scanner-volume distance so that the scanning resolution does not significantly impact the results). Using a threshold of 50 points per cubic meters would lead to a 100% recall and 100% precision. The proposed technique seems to be a good indicator of the presence of shoring, although wrong classification of open volumes could still happen, with both false positive and false negative results. Reducing the risk of false positives would require that the organization of the identified points within the volume be further analyzed (e.g. geometry or color analysis). The risk of false negatives (mainly due to occlusions) could be addressed by developing an algorithm that would estimate the "visible open volume" and thus infer a level of confidence in the result.

Figure 5 Distribution of number of points per cubic meter in the volumes that have shoring and no shoring.

CONCLUSIONS

In this paper, two techniques for recognition of concrete construction secondary and temporary objects in TLS point clouds are proposed. Both of these techniques have shown good potential for recognizing formwork, rebar and shoring using real life data obtained from the Engineering V Buildi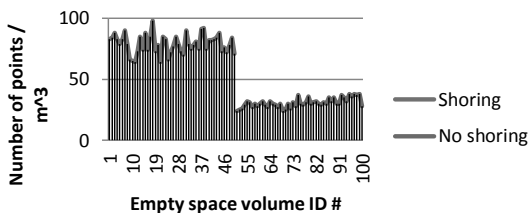ng construction site. The first technique leverages the automated object recognition system of Bosché and Haas (2008). It requires analyzing the level of recognition %$_{recognized}$ within a range of values of τ_D from 10mm to 60mm. Experimental results have shown that it is feasible using this approach to distinguish whether concrete structure objects are in state "built", "formwork" and "rebar". Different settings may be necessary depending on the type of object (columns, walls or floors), as the type of object impacts formwork thickness or rebar cover. The second technique is an application of a simple metric to open spaces defined by design objects and aims at recognizing shoring. The experimental results have shown that it is feasible to differentiate spaces that have shoring and no shoring using this simple metric. It is however acknowledged that the current simple metric, while being a good indicator of the presence of shoring, is not robust and should be extended with additional processing. For example, to ensure the recognition of shoring, algorithms can be developed that would more actively recognize individual shores within the point clouds contained in each open space using techniques such as the Hough transform for 3D edge detection.

Overall, while the seemingly simple approaches (considering the complex foundational software developed to support them) tested in this paper have shown potential, future research is required to explore techniques that could strengthen the recognition of these objects. The integration of color information within the recognition framework has good potential since rebar formwork and shoring typically have colors that are quite different from finished concrete. Analyzing the geometry of points, as investigated by Ishida et al. (2012) for rebar, would further strengthen the recognition performance. Finally, studies remain to be conducted to integrate these results within progress tracking systems (such as the one described in (Turkan et al., 2012)) and measure the resulting improvement in progress tracking.

REFERENCES

Bosché F. and Haas C.T. (2008). "Automated retrieval of 3D CAD model objects in construction range images." Automation in Construction, 17 (4), 499-512.

Bosché F. (2010). "Automated recognition of 3D CAD model objects and calculation of as-built dimensions for dimensional compliance control in construction." Advanced Engineering Informatics, 24 (1), 107-118.

Chen Y. and Medioni G. (1991). "Object modeling by registration of multiple range images." Proceedings of the International IEEE Conference on Robotics and Automation in Sacramento, CA, USA, pp. 2724-2729.

Cheok, G.S., Stone W.C., Lipman R.R., and Witzgall C. (2000). "Ladars for construction assessment and update." Automation in Construction, 9 (5), 463-477.

Golparvar-Fard M., Peña-Mora F., Arboleda C. A., and Lee S. H. (2009). "Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs." Journal of Computing in Civil Engineering, 23 (6), 391-404.

Hurst, M.P. (1983), Formwork, Construction press, Longman Group Limited, New York, NY.

Ishida K., Kano N. and Kimoto K. (2012). "Shape recognition with point clouds in rebars." Gerontechnology Journal, 11(2), 89-96.

Jarkas, A. and Horner, M. (2011). "Revisiting the applicability of learning curve theory to formwork labour productivity." Construction Management and Economics, 29 (5), 483-493.

Lee, S., Oh, S., Hong, M., Choi, J. (2010). "An Investigation of an Image Processing Application for Measurement of a Quantity Estimate of Euro Form Installed in Construction Sites." Proceedings of the IEEE 5th International Conference on Ubiquitous Information Technologies and Applications in Hainan, China, pp. 1-5.

Nunnally, S.W. (2004). Construction Methods and Management, Pearson Prentice Hall, New York, NY.

Peurifoy, R.L. and Oberlender, G.D., (2011). Formwork for concrete structures, McGraw-Hill Professional, New York, NY.

Rusinkiewicz S. and Levoy, M. (2001). "Efficient Variants of the {ICP} Algorithm." Proceedings of the International Conference on 3D Digital Imaging and Modeling (3DIM) in Quebec City, Canada, 2001, pp. 145-152.

Turkan Y., Bosché F., Haas C.T., Haas R. (2012). "Automated Progress Tracking Using 4D Models and 3D Sensing Technologies." Automation in Construction, 22 (1), 414-421.

Wu Y., Kim H., Kim C. and Han S.H. (2010). "Object Recognition in Construction-Site Images Using 3D CAD-Based Filtering." Journal of Computing in Civil Engineering, 24 (1), 56-64.

Development of a System for Automated Schedule Update using 4D Building Information Model and 3D Point Cloud

C. Kim[1], H. Son[2], and C. Kim[3]

[1]Research Assistant, Dept. of Architectural Engineering, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul, Korea 156-756; 82-2-825-5726; 82-2-825-5726; changmin@wm.cau.ac.kr
[2]Research Assistant, Dept. of Architectural Engineering, Chung-Ang University, 221 Heukseok -dong, Dongjak-gu, Seoul, Korea 156-756; 82-2-825-5726; 82-2-825-5726; hjson0908@wm.cau.ac.kr
[3]Associate Professor, Dept. of Architectural Engineering, Chung-Ang University, 221 Heukseok -dong, Dongjak-gu, Seoul, Korea 156-756; 82-2-820-5726; 82-2-812-4150; changwan@cau.ac.kr

ABSTRACT

Correctly generating and updating a schedule are critical to the success of a construction project. However, current schedule update techniques are inaccurate and inefficient because they still rely heavily on manual effort. Although some research proposed an automated schedule update system, it has certain limitations because it uses limited information. This paper describes the development and implementation of a system that automates schedule update using valuable information in the 4D building information model (BIM). The performance of the system was evaluated using a 3D point cloud, obtained during the construction of a project. The developed system improves the schedule update process by integrating 4D BIM with the 3D point cloud. In comparison to current systems, this paper demonstrates a new thinking paradigm in the schedule update process, by using 4D BIM.

INTRODUCTION

Construction managers need to consistently use and periodically update the critical path method (CPM) schedule, in order to successfully complete a construction project (Galloway 2006). Since the 1950s, the CPM has been a valuable tool that has been widely used to schedule and control construction projects (Chanas and Zilnski 2002). By updating a CPM schedule, the manager can calculate the shortest project completion time and identify the critical activities that must be accomplished (Lu et al. 2008). If the project completion time is delayed, construction managers can use CPM to track critical activities, by identifying when the critical-path changed and which activities were delayed (Galloway 2006). Consequently, construction managers who continually update a CPM schedule can control the project and meet its objectives by accelerating delayed activities (Burns et al. 1996).

Although periodic schedule updating is necessary in order to obtain the best results using CPM, it is difficult to update a schedule accurately and efficiently. The schedule update process involves documenting advancements at the construction site, comparing actual progress with planned progress, identifying deviations, and then modifying the schedule (Kiziltas and Akinci 2005; Liu and Shih 2009). Each of the necessary steps in a schedule update requires considerable effort by the construction manager (Bell and McCullouch 1988; Fan et al. 2003). Therefore, in order to assure continuity of the periodic schedule update until construction project completion, accurate and efficient schedule updates are imperative.

Turkan et al. (2012) presented a system to update the schedule using a 3D Computer-Aided Design (CAD) model, schedule information, and 3D point cloud from a 3D laser scanner. According to Turkan et al. (2012), a combination of the 3D CAD model, schedule information, and remote-sensing technology can create an adequate tool to automate schedule updates efficiently. However, the methods of schedule update that were proposed by Turkan et al. (2012) have certain limitations. First, there is no consideration of partial occlusion, full occlusion, or density variation, which almost certainly occurred in the 3D point cloud obtained from the construction site. Second, the updated schedule shows only limited predecessor-successor relationships, which are provided by bar chart. Finally, the only pilot study to support this system was carried out on a linking 3D CAD model using schedule information provided by the authors of the study.

In our previous research (Kim et al. 2013a), we proposed a method that provided information for schedule updates even if partial occlusion, full occlusion, and density variation exist in the 3D point cloud. The key point of this research was to suggest the use of information in the 4D building information model (BIM), in order to provide information to update the schedule. 4D BIM is the building information model that links the 3D BIM with a planned schedule. This 4D BIM is becoming more prevalent in the construction industry.

The current research extends our previous study and develops an automated system that can provide an updated CPM schedule using 4D BIM. The automated schedule update methods are described in the next chapter. Afterward, we present details about the development and implementation of the system. Finally, we summarize our conclusions and propose future research directions.

SCHEDULE UPDATE USING 4D BIM AND 3D POINT CLOUD

This research proposes a process to update the schedule in a specific point in time. The schedule is updated by comparing the as-planned to the as-built at a specific point in time, and then reflecting any deviations between as-planned and as-built. Since 4D BIM links the 3D BIM with the planned schedule, it is possible to generate an as-planned model that extends until the scan date. Data that compare the as-built status of the project to with the as-planned model are represented as a 3D point cloud that is obtained by using a laser scanner. The 3D point cloud obtained from the construction site contains not only structural data, but also various objects, including heavy equipment and construction materials. Thus, in order to obtain the as-built data for the 3D point cloud obtained from the construction site, the 3D point

cloud on the structural components is identified and extracted from the 3D point cloud. After preparation of the as-planned model and the as-built data, the as-planned model and the as-built data are compared, in order to identify the as-built status of each component (e.g. earlier than planned, delayed, and equal to planned). Although the as-built status of each component is recognized, based on the comparison of the as-planned model to the as-built data, there is the problem of full occlusion of the component. In order to solve this problem, an as-built status revision phase is performed. If there is logical inconsistency between the as-built statuses of each component, the as-built status of the component is modified. Then, if there are earlier or delayed components, the schedule is updated by reflecting earlier or delayed components. A detailed flowchart of the proposed process is shown in the Figure 1.
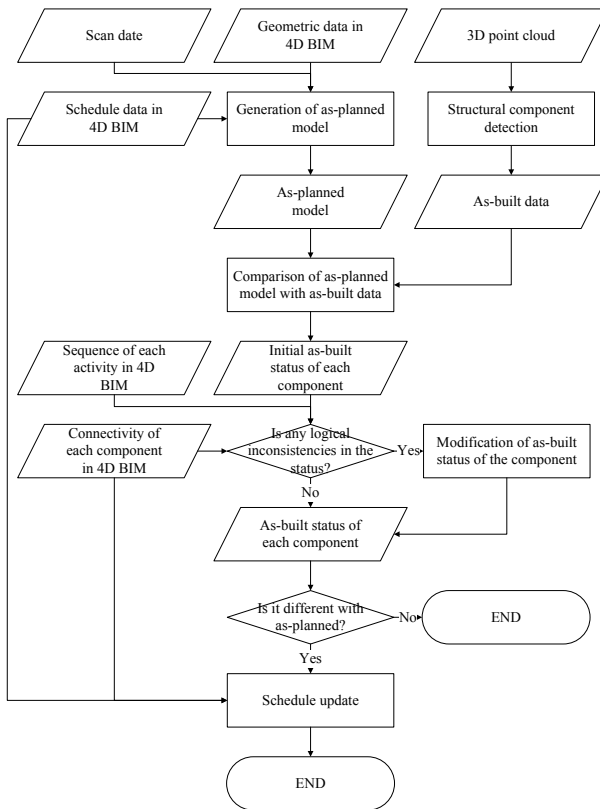

Figure 1. Flowchart of the schedule update

Generation of As-planned Model

In order to use 4D BIM for the schedule update, a user can link activity to the 3D BIM, using BIM solutions, such as Synchro. Most BIM solutions can export 4D

BIM to an industrial foundation classes (IFC) file. IFC is a data model standard that was developed by BuildingSMART International Ltd., as an international information exchange standard. An IFC file contains information on the predefined or user-defined properties of the building's components. Therefore, by extracting the geometric data of each component and the schedule data from the IFC file, an as-planned model can be generated.

The geometric data of each component are used to generate an as-planned model, by linking with schedule data. The geometric data of each component is represented as vertices of surfaces of the component in the IFC file. In order to generate an as-planned model, schedule data should be extracted from the IFC file. There are two data that are related to the schedule in the IFC file. The first is the components assigned to each activity. The components assigned to each activity provide information about components that assign on-going or complete activity until the scan date. The second is the time schedule of each activity. The time schedule of each activity provides the planned time schedule of each activity. Time schedules in the IFC file contain: actual-start, early-start, late-start, schedule-start, actual-finish, early-finish, late-finish, schedule-finish, schedule duration, actual duration, remaining time, free float, total float, and the critical path of each activity. Using the geometric data and schedule data in IFC file, an as-planned model is generated.

Structural Component Detection

An as-built data for update schedule is data about structural components. Structural components have a unique colors; the color of a structural component is different from other objects in the construction site. The 3D point cloud, obtained by using a laser scanner, includes not only position information, but also color information that can be captured by a digital camera. Therefore, the color information of the 3D point cloud and structural component is used to detect the structural component from the 3D point cloud. In order to extract as-built data about structural components from the whole set of the 3D point cloud, a color-model-based structural component detection via a machine learning algorithm proposed by Son et al. (2012) is used. By using the color model of the structural component, as-built data can be accurately extracted from the 3D point cloud obtained from the construction site.

Comparison of As-planned Model with As-built Data

The as-planned model and as-built data obtained from the previous step are compared, in order to identify the as-built status of each component. Because as-planned model and as-built data are three dimensional, the as-planned model can be directly compared to as-built data, by aligning the as-planned model with the as-built data. Alignment of the as-planned model and the as-built data is performed by the 3D registration method proposed by Kim et al. (2013b). This 3D registration method is fully automated align as-planned model with as-built data. Then, matching through identifying constructed components is performed, based on aligned as-planned model and as-built data. Matching of the as-planned model and the as-built data is performed by the method proposed by Kim et al. (2013a). This method is robust to partial

occlusion and density variation of the 3D point cloud. As a result of the matching method, the as-built status of each component is determined.

As-built Status Revision

The as-built status of each component, as determined by the comparison of the as-planned model to the as-built data, may be inaccurate if the 3D point cloud is incomplete. Thus, if there is a logical inconsistency, a revision phase is required to modify the as-built status of each component. The as-built status revision is performed by the method proposed by Kim et al. (2013a). This method makes it possible to modify the inaccurate as-built status of the components, if that inaccuracy is due to an incomplete 3D point cloud. The as-built status of each component is revised, based on the sequence of each activity in 4D BIM and the connectivity between each component in 4D BIM. The sequence of each activity is represented by two activities that have sequence relationships and a relationship between two activities in the IFC file. The connectivity of each component is represented as a list of related and relating components in the IFC file. As a result of the as-built status revision, an accurate as-built status of each component is provided.

Schedule Update

The schedule is updated using the as-built status of each component that was provided in the as-built status revision phase. First of all, if the as-built status of each component is different from the as-planned, the actual duration of an ongoing activity is determined. The actual duration of an ongoing activity is determined by the planned duration based on the deviation between actual and planned. The CPM schedule is updated, based on the determined actual duration of ongoing activities and the sequence relationship that provided by CPM (e.g., start-to-start, start-to-finish, finish-to-start, finish-to-finish). In order to update the CPM schedule, it is necessary to calculate the early start, early finish, late start, late finish, free float, and total float of the successor activities of the ongoing activity. The calculated early start, early finish, late start, late finish, free float, total float, actual duration, and critical path are updated to the time schedule of each activity in the IFC file.

SYSTEM DEVELOPMENT AND IMPLEMENTATION

System Development

Based on the proposed method described previously, the system was developed to update the schedule using 4D BIM and the 3D point cloud. The 4D BIM and the 3D point cloud are automatically processed to the update schedule in the system. Then, the updated schedule is placed in the 4D BIM through the system. The algorithm for data processing was developed on a MATLAB platform. In order to provide a user-friendly application environment, The MATLAB GUI tool is used to design a Graphical User Interface (GUI). The GUI allows the user to both easily import and export data and to choose a specific function just by clicking a button. The

GUI also allows the user to display the list of input files and the status of the schedule update. Figure 2 displays the main menu of GUI.
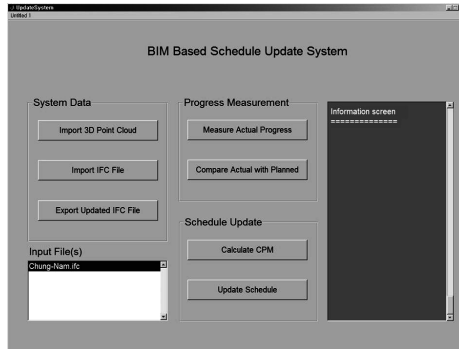
Figure 2. Main menu of GUI

System Implementation

To validate the proposed method and the system developed in this research, a test was conducted on a project which had a 4-story building in South Korea. In this research, the test targeted to the structural work included four activities: ground floor, second floor, third floor, and roof floor. The laser scanner was used to acquire the 3D point cloud from the construction site. Autodesk Revit[TM], developed by Autodesk, Inc., was used to generate the BIM, and Synchro, developed by Synchro Ltd., was used to link the 3D BIM with the schedule data.

Table 1 shows the updated schedule of the project. Ground floor, second floor, and third floor are completed and equal with as-planned. In the case of the roof floor, as-built was earlier than as-planned, so the time schedule of the roof floor was two days less than the as-planned. The critical path was not changed when compared to the as-planned. Because the finish time of critical activity was two days earlier than as-planned, it was expected that the project would finish two days earlier.

Table 1. Updated schedule of the project

Activity Name	As-planned/ As-built	Early start	Early finish	Late start	Late finish	Critical Path
Ground floor	As-planned	100801	100831	100801	100831	O
	As-built	100801	100831	100801	100831	O
Second floor	As-planned	100901	100930	100901	100930	O
	As-built	100901	100930	100901	100930	O
Third floor	As-planned	101001	101031	101001	101031	O
	As-built	101001	101031	101001	101031	O
Roof floor	As-planned	101101	101130	101101	101130	O
	As-built	101101	101128	101101	101128	O

By updating the schedule into the time schedule data in the IFC file, the updated schedule is compatible with commercial scheduling software, such as MS Project and Primavera. Figure 3 shows the import result of the updated schedule into MS project. Likewise, the developed system updates the schedule automatically. Furthermore, because it is compatible with commercial scheduling software, users can use the functions provided by commercial scheduling software, including managing budgets, analyzing workloads, and analytical reporting.
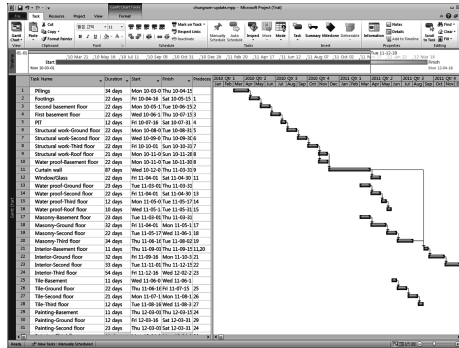
Figure 3. Updated Schedule in MS Project

CONCLUSION

This paper focuses on the development of an automated schedule update system to assist construction managers in controlling the project schedule of the construction phase. The developed system improves the schedule update process, by integrating 4D BIM with the 3D point cloud. Additionally, the system's compatibility with commercial scheduling software makes it easy for the construction manager to control the project schedule. Furthermore, this research unlocks the potential to increase the applicability of the BIM in the construction phase, by using the information in the 4D BIM. In comparison to current systems, this paper demonstrates a new thinking paradigm in the schedule update process, by using 4D BIM.

In current paper, simple and concrete structure was tested to verify the system. However, automated schedule update system is especially required to complex and composite structure. Future research will include verification of the developed system in complex projects that consist of many activities, composite structure, and the various sequence relationships between those activities. In addition, future research will verify the system's use throughout the whole construction phase, including activities such as pile, tile, and curtain wall construction.

ACKNOWLEDGEMENTS

REFERENCES

Bell, L.C., and McCullouch, B.G. (1988). "Bar code applications in construction." *Journal of Construction Engineering and Management*, ASCE, 114(2), 263-278.

Burns, S.A., Liu, L., and Feng, C.-W. (1996). "The LP/IP hybrid method for construction time-cost trade-off analysis." *Construction Management and Economics*, Routledge, 14(3), 265-276.

Chanas, S., and Zielinski, P. (2002). "The computational complexity of the criticality problems in a network with interval activity times." *European Journal of Operational Research*, Elsevier, 136(3), 541-550.

Fan, S.-L., Tserng, H.P., and Wang, M.-T. (2003). "Development of an object-oriented scheduling model for construction projects." *Automation in Construction*, Elsevier, 12(3), 283-302.

Galloway, P. D. (2006). "Comparative study of university courses on critical-path method scheduling." *Journal of Construction Engineering and Management*, ASCE, 132(7), 712-722.

Kim, C., Son, H., and Kim, C. (2013a). "Automated construction progress measurement using a 4D building information model and 3D data." *Automation in Construction*, Elsevier, 31, 75-82.

Kim, C., Son, H., and Kim, C. (2013b). "Fully automated registration of 3D data to a 3D CAD model for project progress monitoring." *Automation in Construction*, Elsevier, in press.

Kiziltas, S. and Akinci, B. (2005). "The need for prompt schedule update by utilizing reality capture the technologies: a case study." *Proceedings of the Construction Research Congress 2005*, San Diego, CA.

Liu, S.-S., and Shih, K.-C. (2009). "Construction rescheduling based on a manufacturing rescheduling framework." *Automation in Construction*, Elsevier, 18(6), 715-723.

Lu, M., Lam, H.-C., and Dai, F. (2008). "Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization." *Automation in Construction*, Elsevier, 17(6), 670-681.

Son, H., Kim, C., and Kim, C. (2012). "Automated color model-based concrete detection in construction-site images by using machine learning algorithms." *Journal of Computing in Civil Engineering*, ASCE, 26(3), 421-433.

Turkan, Y., Bosché, F., Haas, C.T., and Haas, R. (2012). "Automated progress tracking using 4D schedule and 3D sensing technologies." *Automation in Construction*, Elsevier, 22, 414-421.

Fully Automated As-Built 3D Pipeline Segmentation Based on Curvature Computation from Laser-Scanned Data

H. Son[1], C. Kim[2], and C. Kim[3]

[1]Research Assistant, Department of Architectural Engineering, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul, Korea 156-756; PH 82-2-825-5726; FAX 82-2-825-5726; email: hjson0908@wm.cau.ac.kr
[2]Research Assistant, Department of Architectural Engineering, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul, Korea 156-756; PH 82-2-825-5726; FAX 82-2-825-5726; email: changmin@wm.cau.ac.kr
[3]Associate Professor, Department of Architectural Engineering, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul, Korea 156-756; PH 82-2-825-5726; FAX 82-2-825-5726; email: hjson0908@wm.cau.ac.kr

ABSTRACT

There has been a growing demand for the three-dimensional (3D) reconstruction of as-built pipeline. The as-built 3D pipeline reconstruction process consists of measurement of the plant facility, identification of the pipelines, and generation of the 3D pipeline model. Although measurement is now efficiently performed using laser-scanning technology and there has been much progress in 3D pipeline model generation, identification of the pipelines from large and complex sets of laser-scanned data remains a challenging problem. The aim of this study is to propose an as-built 3D pipeline segmentation approach to automatically identify as-built pipelines. The steps of the proposed approach are segmentation of the 3D point cloud, feature extraction based on curvature computation, and pipeline classification. The experiment was performed at an operating plant in order to validate the proposed approach. The experimental result revealed that the proposed method can indeed contribute to the automation of as-built 3D pipeline reconstruction.

Keywords: as-built reconstruction, as-built pipeline; industrial plant; pipeline segmentation; curvature computation

INTRODUCTION

The pipelines play an important role in the operation, maintenance, and expansion of existing chemical, refinery, and power plants (Pottmann et al. 2004). Most of the instruments in plants are connected only by the pipelines, so they act as an intermediary in these instruments' functions (Jin and Lin 2012). Therefore, 3D as-built pipeline models can be used for the maintenance, operation, expansion, or modification of plants (Veldhuis and Vosselman 1998; Ermes 2000). For example, using 3D as-built pipeline models in the planning, expansion, and modification of

existing plants, collisions existing pipelines and planned add-ons can be prevented (Chunmei et al. 2009; Kawashima et al. 2011). During maintenance and operation, 3D as-built pipeline models can be used for the efficient inspection and replacement of parts (Veldhuis and Vosselman 1998).

In practical applications, to reconstruct a 3D as-built model of an entire pipeline, laser scanners are used to measure 3D plants, and then users manually reconstruct a 3D as-built model from the laser-scanned data using several commercial software packages (Chunmei et al. 2009). The laser-scanned data of the existing plant are not only huge, but the pipelines are also intricately entwined like a net (Chunmei et al. 2009; Kawashima et al. 2011). Thus, the users must identify the 3D points corresponding to each pipeline to be modeled in large laser-scanned data sets. To accurately identify the 3D points corresponding to each pipeline, the users need to have some knowledge of the direction and design of the pipelines. In addition, manually identifying each pipeline from the enormous and complex laser-scanned data is a nearly impossible, time-consuming, and labor-intensive process. Nevertheless, previous studies on the reconstruction of 3D as-built pipelines have certain limitations; for example, they still rely on substantial user intervention to identify the 3D points corresponding to each pipeline to be modeled (Kawashima et al. 2011). In addition, although very few research studies have focused on automating this process, they are still limited to only straight portions of an entire pipeline (Rabbani et al. 2006; Bey et al. 2011).

The aim of this study is to propose an as-built 3D pipeline segmentation approach to automatically identify as-built pipelines that occupy large areas of a plant facility from laser-scanned data. The rest of the paper is organized as follows. An overview and details of the proposed approach for as-built 3D pipeline segmentation are provided in Section 2. In Section 3, experimental results are provided. Finally, conclusions and recommendations for future research are given in Section 4.

PIPELINE EXTRACTION BY CURVATURE COMPUTATION

Laser-scanned data acquired from the plant facility can be incomplete because of complex occlusion (Johnson et al. 1997; Rabbani et al. 2006; Masuda and Tanaka 2010), or they can be affected by noise because of the reflective surfaces of the pipelines and other industrial parts. In addition, some applications provide only single scans of data, so part, at most half, of the pipeline surface is acquired (Zheng and Moskal 2012). In this situation, it is challenging to decide where the pipelines are and to even reconstruct the pipelines' geometry and topology.

The proposed as-built 3D pipeline segmentation approach can be divided into three main steps: (1) segmentation of the 3D point cloud at the intersections of the pipelines and other industrial parts, (2) feature extraction based on curvature computation using 30 points for each segment, and (3) pipeline classification to decide whether each segment is pipeline or other industrial part (see Figure 1). The proposed pipeline extraction approach starts with the segmentation of the laser-scanned data. The purpose of this step is to make disjoint the intersections of pipelines and other industrial parts. Next, it is necessary to determine whether each segment belongs to pipelines or other industrial parts. The core idea employed in the

proposed approach is computing curvature at certain points on the objects' surface to decide if it has a cylindrical surface fitting the pipelines' predefined radii. This method requires only one-third of the pipeline surface to compute its radius. Therefore, the proposed approach is applicable even when only single-scan data are available; for example, when only part or, at most, half of the pipeline surface is acquired. Then, based on the results of curvature computation, the segments belong to the pipelines are identified, and all others discarded. The details of the proposed approach are described in the following.
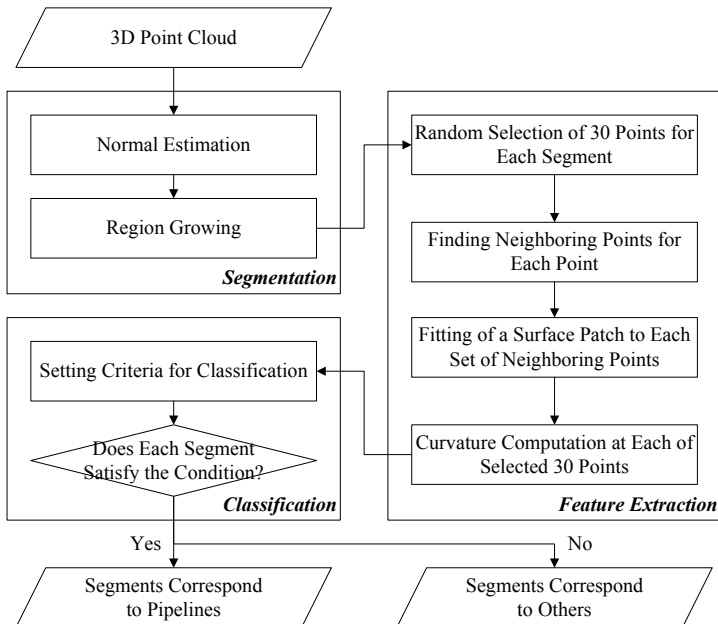
Figure 1. Proposed approach for as-built 3D pipeline segmentation.

Segmentation of 3D Point Cloud

The approach used for segmentation in this study is based on the algorithm proposed in Rabbani et al. (2006) to divide 3D point cloud into a set of disjointed segments. This method uses a criterion based on a combination of surface normal similarity and spatial connectivity, which is defined by Rabbani et al. (2006) as smoothness constraint. This segmentation method is beneficial for detecting smoothly connected areas without encountering over-segmentation problems. Through this process, a given point cloud is divided at the edges and at the areas of high surface normal variation, while preserving the point cloud in each resulting segment locally ensures a smooth surface whose normals do not vary much from each other. We ensure that 3D points that belong to the pipeline can be grouped as one segment even if the pipeline is bent or crossed. For example, when we segment point cloud obtained

from a plant facility, the edges and the areas of high surface normal variation can be intersections of pipelines and other industrial parts, so the resulting segments would belong to one of two groups: pipelines or other industrial parts. Therefore, the segmentation is performed as a pre-processing step for the pipeline segmentation. The segmentation approach consists of normal estimation and region growing, and for more information, please refers to Rabbani et al. (2006).

Feature Extraction Based on Curvature Computation

The surface of the pipeline has cylindrical surface characteristics. In the case of a cylindrical surface, the maximum value of the curvature is equal to the inverse of the radius of the arc. Therefore, this study uses the maximum value of the curvature that provides a general description of the pipelines, which is the radius. The computation of curvature for each segment consists of four steps: selecting 30 points what we called "seed points", finding neighboring points set around the each of the seed points, fitting a local surface patch to the each of neighboring point sets, and computing curvature at the each of the seed points. The computation of curvature for each segment is initiated by randomly selecting 30 seed points. Figure 2(a) shows an example of a segment that belongs to the pipeline and has a 127 mm radius. Figure 2(b) shows 30 seed points (inside the black circle) that were selected by random sampling. By considering 30 points (rather than one point) for each segment, we tried to reduce any potential error that could result from the use of a single point. The red circle in Figure 2(b) shows an example of a randomly chosen seed point' location.

(a)   (b)
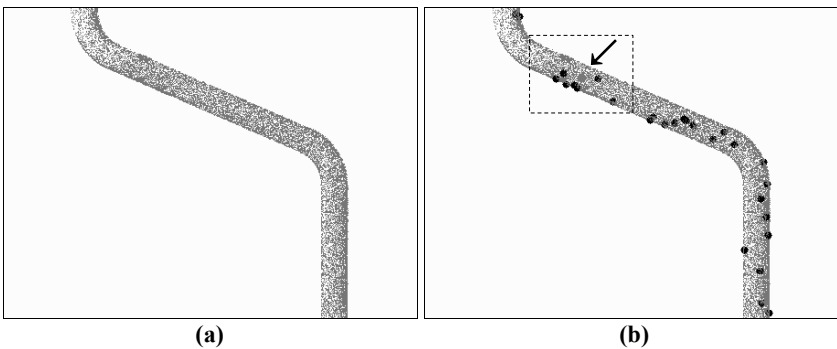
Figure 2. (a) Segmented point cloud that belongs to the pipeline; (b) 30 seed points selected by random sampling.

Then, to compute the maximum value of the curvature for the each of the seed points, its neighboring points are selected based on distance and angle criterion. This process selects only the point of the angle between two normal vectors at a seed point and at a point less than the pre-specified threshold. In this study, when an angle between two normal vectors at a seed point and at a point is less than 120 degrees, the point is considered a neighborhood. Figure 3(a) shows the resulting neighboring points of a seed point. The blue-colored points indicate the resulting neighboring

points of the red-colored seed point. This figure illustrates the magnified portion of the boxed area in Figure 2(b). Figure 3(b) shows the cross-sectional view of Figure 3(a). As shown in Figure 3(b), we could end up with an angle of about 120 degrees.

(a)   (b)
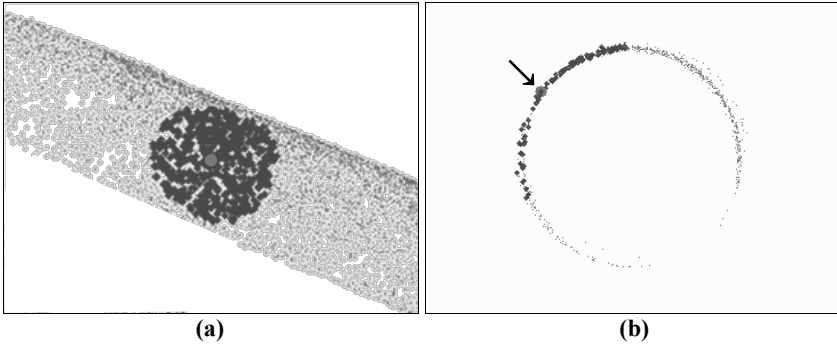
Figure 3. Neighboring points of a seed point: (a) A magnified portion of Figure 2(b); (b) Cross-sectional view.

The next step in the curvature computation is fitting a local surface patch for the neighboring points. In this study, a non-uniform rational B-spline (NURBS) surface was employed as a fitting surface from a given set of the neighboring points (Farin 1997; Piegl and Tiller 1997). With an accurate parameterization in the NURBS surface, it is possible to create a surface patch to best fit the neighboring points (Gálvez and Iglesias 2012). Figure 4(a) shows the result of fitting a local surface patch for the neighboring points (red points). After fitting a local surface patch, surface curvature at the seed point is computed to ascertain maximum value of the curvature. Figure 4(b) illustrates a red circle displaying the maximum curvature at that seed point on the surface patch. These two circles best approximate the curves at the seed point. By inversing the maximum value of the curvature, we can obtain the radius of the smallest circle, which is the radius of the pipeline.

(a)   (b)
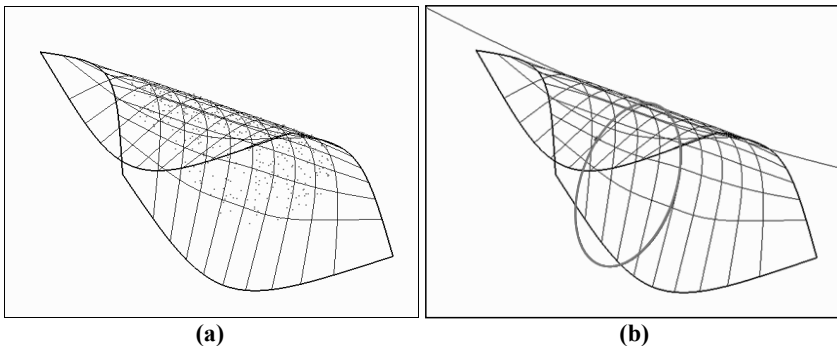
Figure 4. (a) Fitting a local surface patch for the neighboring points (red points); (b) A red circle displays the maximum curvature at the seed point.

Pipeline Classification

For every 30 seed points, we applied the curvature computation process described previously. This study adopted a permissible error of ±5%. For example, if the percent error of a computed radius against a predefined pipeline radius at a seed point is less than 5%, the seed point is considered as part of the pipeline group. If more than or equal to 27 seed points satisfy this condition, the segment is considered to belong to the pipeline group. For example, as stated above, the radius of the pipeline in Figure 6 is 127 mm. For 30 seed points, if at more than or equal to 27 seed points, the percent error of computed radii against 127 mm is less than 5%, and this segment is considered and extracted as part of the pipeline. If not, this segment is considered as part of the "other industrial parts" group, and is discarded.

EXPERIMENTAL RESULTS

The proposed approach described was applied to real experimental data, including cases of partially occluded parts. Laser-scanned data was acquired from the chemical plant located in Yeosu, South Korea. ScanStation C10 by Leica Geosystems was used to acquire 3D point cloud. Figure 5(a) shows the photographic image, and Figure 5(b) shows the laser-scanned data of the test scene. In this scene, there are two types of pipelines, which have radii of 127 mm (a total of eight pipelines) and 177.8 mm (a total of three pipelines).

(a)   (b)
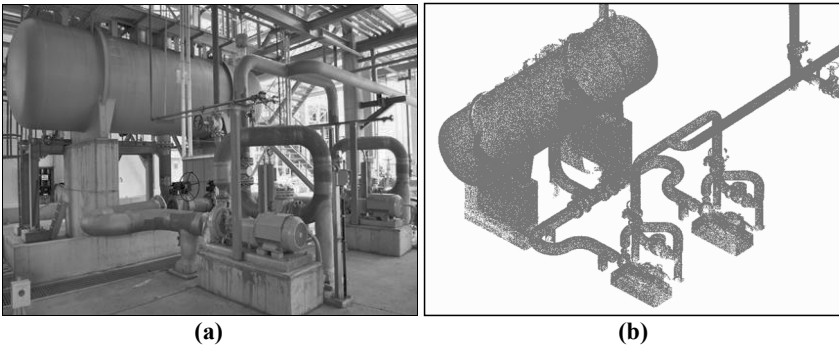
Figure 5. (a) Photographic image; (b) Laser-scanned data.

Figure 6(a) shows the result of segmentation of the point cloud in Figure 5(b). In this figure, different segments are displayed using different colors. This figure shows that the segmentation approach based on smoothness constraint divided point cloud at intersections between pipelines and other industrial parts while a pipeline is preserved as one segment. Figure 6(b) shows the final result of applying the proposed approach for pipeline segmentation. In this figure, blue- and green-colored points indicate the extracted point cloud as the pipelines, which have radii of 127 mm and 177.8 mm, respectively. The proposed pipeline segmentation approach was validated for precision rate. The precision rate shows is calculated as the number of true

pipelines over extracted pipelines. It is observed that the precision rate of the proposed pipeline extraction was 100%. From the experimental result, it can be concluded that the proposed method not only can be used to accurately extract the as-built pipelines but is also useful for the 3D as-built modeling of the pipelines.

(a)   (b)
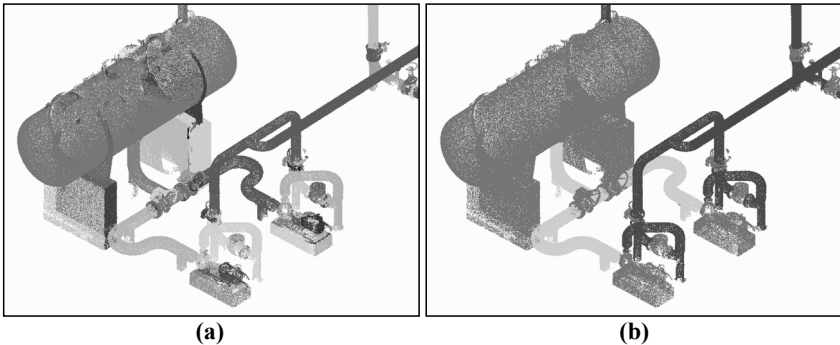
Figure 6. Result of applying the proposed approach: (a) Segmentation of the point cloud; (b) Pipeline Classification.

CONCLUSION

This paper proposes an as-built 3D pipeline segmentation approach to automatically identify as-built pipelines that occupy a large portion of the as-built plant facility. The segmentation of the point clouds is performed based on smoothness constraint. The segments of laser-scanned data are then classified as either pipeline or other industrial parts based on the curvature computation using 30 randomly chosen points of the surface of each segment. The feasibility of the proposed approach was validated in an experiment using real laser-scanned data obtained from an operating industrial plant. The results demonstrated that the proposed approach can successfully segment the as-built 3D pipelines with other industrial parts present. In this process, each pipeline' radius is determined, therefore, it can be extended for the as-built 3D pipeline reconstruction. The proposed approach is advantageous, because it segments and identifies as-built pipelines even when the pipelines' surface is partially invisible due to partial occlusion by other objects. Thus, it could be successfully incorporated into applications when only single-scan data are available, such as the detection and identification of underground pipelines. Nevertheless, the proposed approach was validated and tested for only two types of pipelines with a radius of 127 mm and 177.8 mm. For the purposes of practical use, future research should focus on the validation on all types of pipelines and improving the proposed approach.

ACKNOWLEDGEMENTS

REFERENCES

Bey, A., Chaine, R., Marc, R., Thibault, G., and Akkouche, S. (2011). "Reconstruction of consistent 3D CAD models from point cloud data using a priori CAD models." *Proc. Int. Society for Photogrammetry and Remote Sensing Laser Scanning 2011*, Calgary, Canada.

Chunmei, H., Qiusheng, Z., Wentao, Y., and Jingtang, Z. (2009). "Research on 3D reconstruction of petrochemical enterprise pipeline and its collision detection based on 3D-scanning technology." *Proc. Int. Forum on Computer Science-Technology and Applications*, Chongqing, China.

Ermes, P. (2000). "Constraints in CAD models for reverse engineering using photogrammetry." *Proc. Int. Archives of the Photogrammetry and Remote Sensing*, Amsterdam, Netherlands.

Farin, G. (1997). *Curves and surfaces for computer-aided geometric design: A practical guide*, Academic Press, New York, NY.

Gálvez, A., and Tglesias, A. (2012). "Particle swarm optimization for non-uniform rational B-spline surface reconstruction from clouds of 3D data points." *Information Sciences*, 192(1), 174–192.

Jin, X., and Lin, C. (2012). "Research on the whole process cost control methods of the chemical piping engineering." *Advanced Materials Research*, 383–390, 4286–4293.

Johnson, A. E., Hoffman, R., Osborn, J., Hebert, M. (1997). "A system for semi-automatic modeling of complex environments." *Proc. Int. Conf. on Recent Advances in 3-D Digital Imaging and Modeling*, Ottawa, Canada.

Kawashima, K., Kanai, S., and Date, H. (2011). "Automatic recognition of a piping system from large-scale terrestrial laser scan data." *Proc. Int. Society for Photogrammetry and Remote Sensing Laser Scanning 2011*, Calgary, Canada.

Masuda, H., Tanaka, I. (2010). "As-built 3D modeling of large facilities based on interactive feature editing." *Computer-Aided Design and Applications*, 7(3), 349–360.

Piegl, L., and Tiller, W. (1997). *The NURBS book*, Springer-Verlag, New York, NY.

Pottmann, H., Hofer, M., Odehnal, B., and Wallner, J. (2004). "Line geometry for 3D shape understanding and reconstruction." *Proc., European Conf. on Computer Vision*, Prague, Czech Republic.

Rabbani, T., Van den Heuvel, F., and Vosselman, G. (2006). "Segmentation of point clouds using smoothness constraint." *Proc. ISPRS Commission V Symp. on Image Engineering and Vision Metrology*, Dresden, Germany.

Veldhuis, H., and Vosselman, G. (1998). "The 3D reconstruction of straight and curved pipes using digital line photogrammetry." *ISPRS Journal of Photogrammetry and Remote Sensing*, 53(1), 6–16.

Zheng, G., and Moskal, L. M. (2012). "Spatial variability of terrestrial laser scanning based leaf area index." International Journal of Applied Earth Observation and Geoinformation, 19, 226–237.

Grand Challenges in Simulation for the Architecture, Engineering, Construction and Facility Management Industry

SangHyun Lee[1], Amir Behzadan[2], Amr Kandil[3], and Yasser Mohamed[4]

[1]Assistant Professor, Department of Civil and Environmental Engineering, University of Michigan, 2340 GGB, 2350 Hayward Street, Ann Arbor, MI 48109; PH (734) 764-9420; FAX (734) 764-4292; email: shdpm@umich.edu
[2] Wharton Smith Faculty Fellow and Assistant Professor, Department of Civil, Environmental, and Construction Engineering, University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816-2450; PH (407) 823-2480; FAX (407) 823-3315; email: amir.behzadan@ucf.edu
[3] Assistant Professor, School of Civil and Environmental Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907; PH (765) 494-2246; FAX (765) 494-0644; email: akandil@purdue.edu
[4] Associate Professor, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, T6G 2W2; PH (780) 492-1321; FAX (780) 492-0249; email: yaly@ualberta.ca

ABSTRACT

Today's Architecture, Engineering, Construction and Facility Management (AEC/FM) industry has to deal with complex obstacles (e.g., aging infrastructure, the protection of the natural environment, the need for resilient infrastructure). Among several technical fields that have been adopted and advanced, computer simulation has been widely researched and practiced for the effective delivery and maintenance of capital projects. Considering the unprecedented problems of today's infrastructure and the rapid advancements in simulation research, it is very timely to investigate what grand challenges exist in simulation that if properly addressed, can help create a more sustainable and resilient engineering of infrastructure lifecycle. To this end, this paper aims to identify grand challenges in simulation by presenting major areas of interest that can benefit the AEC/FM domain and discussing three specific areas identified as challenges in simulation: 1) realistic simulation modeling; 2) applicability of simulation models to the industry; and 3) academic and educational obstacles. Specific foci are the knowledge gaps in these major areas, and the current efforts to mitigate these gaps based on extensive literature reviews. This paper also proposes the next steps for addressing these challenges based on the findings herein.

INTRODUCTION

Following a major infrastructure expansion during the second half of the 20th century, today's AEC/FM industry has to deal with the technical issue arising from

the aging U.S. infrastructure while thinking about protecting the natural environment (i.e., sustainability). Among several technical fields that have been adopted and advanced to address these challenges, computer simulation has been widely researched and practiced for the effective delivery and maintenance of capital projects. Computer simulation has thus demonstrated its usefulness in aiding AEC/FM engineers in designing, constructing, and operating infrastructure systems. However, taking into account the unprecedented challenges faced by today's infrastructure (e.g., aging, the need for resilience and sustainability), as well as the rapid advancement of new simulation techniques, it is very timely to investigate what grand challenges exist in simulation. This effort can transform the current practice of infrastructure lifecycle engineering into a more sustainable and resilient process. In this context, the Visualization, Information Modeling, and Simulation (VIMS) committee—one of the technical committees under the ASCE's Technical Council on Computing and Information Technology—recently initiated an expert taskforce to study and define grand challenges in VIMS. As its first effort, VIMS committee members had a focus-group-like discussion at their 2012 meeting, held in Clearwater Beach, FL. At this meeting, the committee identified several areas where challenges exist in VIMS, such as: 1) generating realistic simulation models (e.g., incorporating human behavior and addressing discrepancies between simulation models and reality); 2) improving the applicability of simulation models to the industry (e.g., accreditation and adoption of simulation models and real-world validation); and 3) addressing academic and educational obstacles (e.g., integration of simulation into engineering curricula and simulation for interdisciplinary research). This paper reports the knowledge gaps in the above areas and the current efforts to mitigate these gaps based on extensive literature reviews in order to continue to discuss grand challenges in simulation with potential benefits to the AEC/FM domain.

GENERATING REALISTIC SIMULATION MODELS

As computer simulation has been widely adopted, the need to incorporate more details for realistic model generation has become more imperative. Two identified topics that pose a great challenge in this area are incorporation of human and social behavior and realistic representation of simulation models that are responsive to real-world data.

Incorporation of Human Behavior

Throughout its lifecycle, a built environment has interactions with a large number of people. From the production perspective, the built environment is a physical product of integrating human labors of various forms, and from the use perspective (i.e., operation) it provides the setting for various human activities. Due to such a rich interaction, human factors greatly affect both the production and the use of built environments (Wickens and Hollands 2000). Therefore, the understanding of human factors is invaluable for sound decision making in the design, engineering, construction, and operation of the built environment. However, studying human factors is not easy and straightforward, mainly because of the difficulty in conducting human-centered experiments. One way to study human factors is to use computer

simulation, which enables an end user to conduct experiments with a model as many times as they wish, and provides observations on the model's behavior (i.e., response) to different inputs (i.e., scenarios). Classes of research topics in this direction include building emergency evacuation simulation (Shi et al. 2009; Gwynne et al. 2005; Galea 2003; Proulx and Richardson 2002; Chu et al. 2011; An et al 2009), design performance simulation (Dijkstra 2008; Willis et al. 2004; Dijkstra and Timmermans 2002; Yan and Kalay 2006), occupant behavior for energy simulation (Clevenger and Haymaker 2006; Rijal et al. 2008; Azar and Menassa 2011; Hoes et al. 2009; Chen et al 2012; Anderson et al. 2012; Peschiera et al. 2010), construction worker behavior (e.g., productivity and absence) (Levitt et al 1999; Jin and Levitt 1996; Watkins et al. 2009; Ahn et al. 2012; Ahn and Lee 2011), and project organizations simulation (Comu et al. 2011; Chinowsky et al 2011; Chinowsky and Songer 2011). These topics are researched to increase the understanding of human behavior interacting with built environments and improve the realism of simulation models. For example, in the study of project organizations' performance, the focus has been placed upon the involvement of numerous people from diverse organizations (e.g., owners, designers, engineers, construction managers, contractors, and trade unions); this involvement can greatly affect project organizations' performance. At the organizational level, social interaction between parties affects a project's performance (Comu et al. 2011; Chinowsky 2011). At the personal level, interactions between project participants affect a project's performance. In turn, the performance of the entire project is affected (Chinowsky and Songer 2011). From this review, a number of new challenges can be identified. For instance, a comprehensive simulation framework is needed to integrate segmented human behavior simulation models and their other interactive models. For example, energy simulation can use DOE-2 to estimate energy consumption, but if the simulation needs to incorporate varying occupant energy consumption behavior, an occupant behavior model has to be integrated for a comprehensive analysis. In addition, modeling human behavior needs domain-specific knowledge; otherwise a model could remain distant from reality. For example, applying absence culture to construction workers' behavior can be valid only when the characteristics of construction projects are fully considered (e.g., workers move to different sites while existing theory about absence culture has been made for those who stay in their company indefinitely, such as nurses or bus drivers).

Realistic Simulation Model Responsive to Real-World Changes
Effective simulation-based decision making requires that all model parameters and variables are defined as real-time functions of ongoing project performance, and are modified with time if and when necessary, so that ultimately the simulation model is completely adaptable to the latest conditions on the ground (Xie and Abourizk 2011; Akhavian and Behzadan 2012). However, many existing simulation tools that have been developed for use by the AEC/FM industry are structured to create only intuitive means to study a facility or a project during early planning and design stages. As such, the resulting simulation models are built upon (often subjective) assumptions about major model parameters and performance indicators, which were made during the infancy stages of a project. Thus, a major research challenge is to generate realistic simulation models that are responsive to changes in the real engineering system

during the execution phase. Such a comprehensive methodology for incorporating real time field data would release modellers from relying on data from previous projects, expert judgments, and rigid assumptions. Collecting reliable and accurate data from a highly dynamic environment such as a construction jobsite is almost impossible using traditional (manual) methods. Thanks to the latest technology advancements, several automated data collection and processing techniques have recently been introduced and used by researchers within the AEC/FM domain, including global positioning systems (GPS), radio frequency identification (RFID), ultra wide-band (UWB), and 3D laser scanners that assist in tracking objects (e.g., equipment, material, personnel), detecting unsafe work zones, or measuring productivity (Oloufa et al. 2003; Caldas et al. 2006; Song et al. 2006; Ergen et al. 2007; Jang and Skibniewski 2007; Teizer et al. 2008; Behzadan et al. 2008; Grau and Caldas 2009; Khoury and Kamat 2009; Brilakis et al. 2011; Andoh et al. 2012; Han et al. 2012). While an autonomous data collection system provides higher reliability and accuracy, it should also be able to process and use the collected data in (near) real time, thereby facilitating short-term decision making, planning, and control. For this purpose, the efficiency and robustness of the data collection process turns out to be as important as data reliability and accuracy. Once data about project entities is collected, it needs to be transformed into meaningful process information, and ultimately into operational knowledge. Without putting field data into proper context, it is impossible to generate simulation models that describe the real operations as they occur. Existing single-modal remote sensing and information extraction methods can be easily misled by incomplete detection. For example, most vision-based methods suffer from occlusions, and methods that rely on preinstalled infrastructure cannot detect events that constantly evolve (e.g., road construction). These issues can create problems when detecting all operational details as they occur, and thus highlight the need for a systematic approach to capturing and fusing data from multiple sources (Pradhan and Akinci 2012). The challenge is to find the "just right" volume of reliable data that represents a proper level of detail. Very often, captured data is of tremendously large volume and contains high noise ratio (Razavi and Haas 2012), such that data cleaning and analysis takes a considerable amount of time. For instance, to study the performance of a few workers, data collected using a video camera may take a large amount of time to process yet not yield the desired accuracy or robustness (Yang et al. 2010). Even if data can be collected in a timely manner, the desirable setting and the expected spatio-temporal relations between project entities may be adversely affected by unforeseen conditions. Thus, one has to also make sure that raw field data is rich enough to reflect such special cases. To this end, the main research challenge is to determine how much data, and with what resolution, is enough to represent the real world system with an acceptable level of detail in the least possible time.

IMPROVING THE APPLICABILITY OF SIMULATION MODELS TO THE INDUSTRY

Despite advancements in computer simulation, its usage in the AEC/FM industry has been limited. Two important topics for a wider adoption by the industry

have been identified: how to increase the credibility of simulation models, and how to increase user confidence with validation and verification.

Credibility and Adoption of Simulation Models

The adoption of innovative and new technologies by the construction industry is not a trivial decision. A workshop sponsored by NSF and CII identified key factors that need to be fulfilled in order for a new technology to be accepted. These factors are 1) new technology compatibility with the "company culture"; 2) availability of a champion who can demonstrate the advantages of the new technology and push its adoption; and 3) realization that the new technology yields significant benefits relative to its overall cost (Oberlender 1997). It should be noted that the term "new" is relative to the industry as a whole, to a sector or specialization of it, to a particular company, or even to a department or work team within a company. Simulation is a technology that has been around and in use by different industries for decades. However, it may still be new for a particular construction firm relative to its common practice. Hence, the adoption of simulation models by the construction industry is a strategic decision. A few researchers have reported on this topic, and their successful adoption stories all seem to align with the above factors. For example, Halpin and Martinez (1999) identified the exact same factors as keys to the successful adoption of simulation technology by Dragados y Construcciones. AbouRizk et al. (2011) and AbouRizk (2010) also listed a number of successful cases of simulation adoption where the same factors were present. They added another key factor—"trust" that develops over time between a company and research team/simulation developer. The strategic level factors identified above are irrelevant to the type of simulation modeling approach or tools used. At a more detailed and operational level, many factors have been identified as important aspects of a simulation study to facilitate better credibility, acceptance and adoption of simulation results. Some of the factors sought by construction researchers are: better communication of model behaviors and output through advanced visualization (Kamat and Martinez 2003), faster model development cycle through easier-to-use tools (AbouRizk and Hajjar 2002), and more efficient construction operation knowledge acquisition (AbouRizk et al. 2011). Simulation researchers on the other hand have identified a much wider spectrum of factors that simulation studies have to fulfill to achieve success. For example, Robinson and Pidd (1998) identified 338 factors classified along 19 dimensions based on interviews with modellers and clients of simulation studies. It should be noted that software tools represent only one of these dimensions, while most of the dimensions relate to the development process and modeller capabilities. A key observation on discussions related to the acceptance and adoption of simulation within the construction domain and the simulation community at large (Robinson 2004), is that they are mostly based on experience rather than empirical evidence. There is a need in the construction domain to report cases of simulation studies, with various levels of success, to facilitate analysis and the development of a framework for the successful integration of simulation studies within construction decision-making processes. Such cases must include a description of the management process from start to end, and not only model algorithms and output analysis.

Verification, Validation, and Confidence

Verification, validation, and testing of simulation models are ways to build model credibility, which in turn raises confidence in model output and leads to adoption and implementation. However, simulation models, by design, are built to test situations and scenarios that are not present in the real world. Therefore, validating them against real-world data of the exact same situation they will be used for is, to an extent, an oxymoron. Alternatively, we try to test the model against historical data from a situation "close enough" to our intended use. This concept is discussed by several authors and they all seem to agree that validation is both an art and a science (Balci 1998). Robinson (2004) states that: "It is not possible to prove that a model is absolutely correct." In fact most verification and validation efforts and tests, including statistical ones, try to check if a model is incorrect. The more they fail at proving that a model is incorrect, the more our confidence in that model is increased (Sergant 2000). Another key aspect of the validation and credibility of a simulation model is their unification with the purpose and intended use of the model where a model is considered an accurate enough representation of a system for a particular study objective but cannot be generalized beyond that objective (Law 2009). Projecting these principles on construction simulation studies complicates things even further and calls for a special framework for assessing the validity and credibility of construction simulation models. The uniqueness of construction projects makes it difficult to validate a model against real world data. This challenge strengthens the need for full documentation and reporting of such simulation case studies. The documentation should provide access to datasets to enable the validation of new models. In addition, efforts made in other domains with similarities to construction are worth considering. For example, building stochastic simulation models of combat scenarios in defense modeling and simulation is similar to a construction project in that a combat scenario is unique too and thus, validating a computer simulation model with limited actual data becomes a challenge. To this end, Champagne and Hill (2009) proposed a statistical approach (i.e., a bootstrapping technique) to overcome this issue; their approach could be considered for construction projects as well.

ADDRESSING ACADEMIC AND EDUCATIONAL OBSTACLES

Another area that has been called for is how to maximize the benefits of simulation in academia. Two representative topics are the integration of simulation into engineering curricula and effectively conducting interdisciplinary research.

Integration of Simulation into Engineering Curricula

Integrating simulation technologies into engineering curricula can greatly contribute to students' learning, enabling them to model real-world problems and experiment with what-if scenarios. However, this integration also poses significant challenges such as the need for computational resources to be constantly updated. The computational resources that most schools focus on are the ones that get frequently used and are needed to train students to meet workplace demands. Therefore, the adoption of certain computational technologies has been gradual due to the fact that they have been slowly adapted by the industry (Sampio et al. 2009). The availability

of computational resources is directly linked to the need for manual interaction with models in a shared workspace that characterizes collaborative learning (Dong et. al. 2012). This creates a challenge that is exasperated by the misconceptions that model developers often have about how actual systems operate. In order to avoid this major pitfall, models need to be validated in a process that recognizes and corrects modeling errors and determines how accurately a simulation model represents the real world. This process requires the involvement of practitioners and decision makers who have intimate knowledge of the actual system to insure its practical relevance (Dong et al. 2012; Sampio et al. 2009). Computational resources are extrinsic to the simulation model itself. An intrinsic challenge, then, is the extent to which students understand the structure of the simulation. This understanding greatly impacts how they learn. The students' understanding of simulation structure is directly impacted by the transparency of the underlying simulation structure made available to them. This transparency falls into one of three categories: (1) "black box" models that keep the structure of the model hidden from users, generating only model behaviors; (2) "cloudy box" models where the model structure is partially visible by [for example] providing conceptual model structures but not providing their formalizations; and (3) "glass box" models that are completely transparent such that the model structures that drive their behavior are visible to and understandable by the user. The factor that should be initially determined while ascertaining the degree of transparency is how much time and effort is needed to develop students' background and understanding of the model structure. Naturally, more transparent models would require additional time and effort (Ford 2010). To address this challenge, several issues should be considered in model selection, including: (1) the suitability of different simulations to specific learning contexts; (2) the pedagogical approaches that could be used to support learning outcomes; and (3) the validity of using a chosen game or simulation (Sampio et al. 2009; Wall and Ahmed 2008). This creates a need for research into how simulation can be integrated into the learning process (Wall and Ahmed 2008).

Finding Effective Ways to Conduct Interdisciplinary Research

In order to facilitate interdisciplinary research, simulation methodologies must allow researchers to: (1) integrate multiple modeling approaches that are suitable for different components or research disciplines; and (2) represent model outputs in a unified and readily accessible visualization method. Many real-world systems are composed of both feedback and sequential processes. These systems also include combinations of context and operation-level parameters, as well as discrete and continuous variables (Lee et al. 2009). Despite this reality, most simulation platforms do not permit the integration of methods that allow for the accurate representation of real-world processes. Therefore, system modelers are in many cases forced to make simplifying assumptions that reduce the fidelity of the simulation (Alvanchi et. al. 2010). One of the approaches to solving this problem is to develop custom software that allows for the integration of two or more simulation methods. These custom efforts, however, do not provide sufficient flexibility for modeling problems from different disciplines (Alvanchi et al. 2010). The integration of different simulation methods presents challenges such as the communication of data between different model components. Additionally, the main purpose of this integration is to reflect the

effects of the latest changes in different parts of the system on the operation of other parts. This could potentially lead to a large simulation processing time overhead required for updating different system components (Alvanchi et. al. 2009). Another problem that hinders the widespread application of simulation methods is the difficulty faced by stakeholders in interpreting model outputs. Therefore, computer visualization methods have been proposed to generate more intuitive and easier to interpret output (Zhang et al. 2012). In the case of interdisciplinary research, the amount of programming needed to develop visualization tools (e.g., graphic engines) supporting those simulation methods is significant. Unfortunately, there is a very limited availability of ready-to-use generic graphics engines that could be used with simulation methodologies across disciplines (Zhang et al. 2012). There have been some attempts to address these limitations through distributed and interactive simulation design methodologies. The two notable efforts include the development of Distributed Interactive Simulation (DIS) (IEEE standard 1278) and the creation of High Level Architecture (HLA) (IEEE Standard 1516), which was a successor to DIS. Both DIS and HLA create architectures for interconnecting separate simulations (Zhang et al. 2012). HLA is popular in both the military and civilian domains that need high reusability and interoperability and its goal is to promote interoperability between simulation system, which would facilitate the reuse of models in different contexts and domains. This reuse ultimately reduces the time and cost needed to create new simulation environments (AbouRizk 2010). Despite such efforts and contributions made to enhance the use of simulation techniques in interdisciplinary research (AbouRizk 2010; Alvanchi et al; 2009), there is still a strong need for research in the area, specifically in developing simulation methodologies based on HLA. These simulation systems can then be integrated across disciplines and problem domains to create simulation models that accurately represent the real-world situation.

CONCLUSIONS

We discussed the details about three major challenging areas that were identified in the 2012 ASCE TCCIT VIMS committee meeting: realistic simulation model generation, wider applicability of simulation models to the industry, and academic and educational challenges. Based on these discussions, major research challenges were identified that included but were not limited to fusion of new multi-modal data-capture technologies, incorporation of newly introduced social science perspectives, and the need for interdisciplinary research. However, there are also challenges that have been lingering since the introduction of simulation to AEC/FM, such as model credibility, real-world validation, and integration of simulation topics into engineering curricula. Based on these observations, we concluded that the next logical steps would be to extend these areas and provide more inclusive solutions, and then to divide them into two major categories—new and existing challenges—identified by unique set of characteristics (e.g., why and how new challenges are different from existing ones, and what situations have been changed in the case of existing challenges). Consequently, a thorough discussion is needed on how our evolving knowledge on simulation, as well as the advancements in simulation technologies can address these challenges.

REFERENCES

AbouRizk, S. (2010). "The role of simulation in construction engineering and management." *J. Constr. Eng. Manage.*, 136(10), 1140–1153.

AbouRizk, S., and Hajjar, D., "Unified Modeling Methodology for Construction Simulation" *J. Constr. Eng. Manage.*, 128(2), 174–185.

AbouRizk, S., Halpin, D., Mohamed, Y., and Hermann, U. (2011). "Research in Modeling and Simulation for Improving Construction Engineering Operations." *J. Constr. Eng. Manage.* 137, SPECIAL ISSUE: Construction Engineering: Opportunity and Vision for Education, Practice, and Research, 843-852.

Ahn, S. and Lee, S. (2011). "Absence Norm in Construction Project Organization and Its Implication in Human Resource Management: Using Agent-Based Modeling." *Proceeding of the Winter Simulation Conference*, Institute of Electrical and Electronics Engineers (IEEE), Piscataway, NJ.

Ahn, S., Lee, S., and Park, M. (2012). "Exploring Absenteeism Control Policies with Awareness of the Effect of Group Norms on Absence Behavior, Using Agent-Based Modeling." *2012 ASCE Construction Research Congress*, West Lafayette, IN.

Akhavian R., and Behzadan A.H. (2012). "An integrated data collection and analysis framework for remote monitoring and planning of construction operations", *Journal of Advanced Engineering Informatics*, Elsevier, 26(4), 749-761.

Alvanchi, A., Lee, S., and AbouRizk, S. M. (2009). "Modeling Framework and Architecture for Hybrid System Dynamics and Discrete Event Simulation for Construction." *Computer-Aided Civil and Infrastructure Engineering*, Blackwell, 26 (2), 77-91.

An, S., Cui, N., and Wang, J. (2009, July). "Modeling and Simulating Household Evacuation Behaviors for Evacuation Time Estimates". *In International Conference on Transportation Engineering 2009,* ASCE, 4134-4140.

Anderson, K., Lee, S. and Menassa, C. (2012). "Effect of Social Network Type on Building Occupant Energy Use." *BuildSys 2012 Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, ACM, New York, 17-24.

Andoh A.R., Su X., and Cai H. (2012). "A framework of RFID and GPS for tracking construction site dynamics", *Proceedings of the ASCE Construction Research Congress (CRC)*, West Lafayette, IN.

Azar, E., and Menassa, C. C. (2011). "Agent-Based Modeling of Occupants' Impact on Energy Use in Commercial Buildings". *Journal of Computing in Civil Engineering*.

Balci, O. (1998) "Verification, Validation, and Accreditation" *In proceedings of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson, and M.S. Manivannan (eds.), Vol. 1, 41-48

Behzadan A.H., Aziz Z., Anumba C.J., and Kamat V.R. (2008). "Ubiquitous location tracking for context-specific information delivery on construction sites", *Journal of Automation in Construction*, Elsevier, 17(6) 737-748.

Brilakis I., Park M.W., and Jog G. (2011). "Automated vision tracking of project related entities", *Journal of Advanced Engineering Informatics*, Elsevier, 25(4) 713-724.

Caldas C.H., Grau D.T., and Haas C.T. (2006). "Using global positioning system to improve materials-locating processes on industrial projects", *Journal of Construction Engineering and Management*, ASCE, 132(7), 741-749.

Champagne, L. E., and Hill, R. R. (2009). "A Simulation Validation Method Based on Bootstrapping Applied to an Agent-based Simulation of the Bay of Biscay Historical Scenario", *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 6 (4), 201-212.

Chen, J., Taylor, J. and Wei, H. H. (2012). "Modeling Building Occupant Network Energy Consumption Decision-making: The Interplay between Network Structure and Conservation," *Energy and Buildings*, 47, 515-524.

Chinowsky, P. S. (2011). "A network and culture perspective on organizational management." in Chinowsky, P. S. and Songer, T. (eds) Organization Management in Construction, Taylor and Francis, London.

Chinowsky, P., Taylor, J., and Di Marco, M. (2011). "Project Network Interdependency Alignment: A New Approach to Assessing Project Effectiveness," *Journal of Management in Engineering*, ASCE, 27(3): 170-178.

Chu, M. L., Pan, X., and Law, K. (2011, June). "Incorporating Social Behaviors in Egress Simulation". *In Proceedings of 2011 Computing in Civil Engineering Workshop,* 19-22.

Clevenger, C. M., and Haymaker, J. (2006). "The impact of the building occupant on energy modeling simulations". *In Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montreal, Canada, 1-10.

Comu, S., Unsal, H., and Taylor, J. (2011). "The Dual Impact of Cultural and Linguistic Diversity on Project Network Performance," *Journal of Management in Engineering*, ASCE, 27(3): 179-187.

Dijkstra, J., and Timmermans, H. (2002). "Towards a multi-agent model for visualizing simulated user behavior to support the assessment of design performance", *Automation in construction*, 11(2), 135-145.

Dijkstra, J. (2008, December). "An Agent Architecture for Visualizing Simulated Human Behavior to Support the Assessment of Design Performance", *In Computational Intelligence for Modelling Control & Automation*, 2008 International Conference on IEEE, 808-813.

Dong, S., Behzadan, A., Chen, F., and Kamat, V., R. (2012) " Collaborative visualization of engineering processes using tabletop augmented reality" *Advances in Engineering Software*, Elsevier, 55, 45-55.

Ergen E., Akinci B., East B., and Kirby J. (2007) "Tracking components and maintenance history within a facility utilizing radio frequency identification technology" *Journal of Computing in Civil Engineering*, ASCE, 21(1), 11–20.

Ford, D. (2010) "Integrating Simulation into Research and Teaching of Construction Engineering and Management: Reflections on Experience" *Proceedings of the 2010 Winter Simulation Conference*, Baltimore, MD, December 5-8, 2010, 3049-3055.

Galea, E., (ed.). (2003). "Pedestrian and Evacuation Dynamics" *Proceedings of 2nd International Conference on Pedestrian and Evacuation Dynamics*, CMC Press, London.

Grau D.T, and Caldas C.H. (2009). "Methodology for automating the identification and localization of construction components on industrial projects" *Journal of Computing in Civil Engineering*, ASCE, 23(1), 3-13.

Gwynne, S., Galea, E. R., Owen, M., and Lawrence, P. J. (2005). "The Introduction of Social Adaptation within Evacuation Modeling." *Fire and Materials*, 2006(30), 285-309.

Halpin, Daniel W. and Martinez, Luis-Henrique (1999). "Real world applications of construction process simulation." *In Proceedings of the 31st conference on Winter simulation: Simulation---a bridge to the future*, Phillip A. Farrington, Harriet Black Nembhard, Gerald W. Evans, and David T. Sturrock (eds.), Vol. 2. 956-962.

Han S., Lee S., and Peña-Mora F. (2012). "A machine-learning classification approach to automatic detection of workers' actions for behavior-based safety analysis" *Proceedings of ASCE International Workshop on Computing in Civil Engineering*, Clearwater, FL.

Hoes, P., Hensen, J. L. M., Loomans, M. G. L. C., De Vries, B., and Bourgeois, D. (2009). "User behavior in whole building simulation." *Energy and Buildings*, 41(3), 295-302.

Jang W.S., and Skibniewski M.J. (2007) "Wireless sensor technologies for automated tracking and monitoring of construction materials utilizing Zigbee networks" *Proceedings of the ASCE Construction Research Congress (CRC)*, Grand Bahamas Island.

Jin, Y., and Levitt, R. E. (1996). "The virtual design team: A computational model of project organizations". *Computational & Mathematical Organization Theory*, 2(3), 171-195.

Kamat, Vineet R., and Martinez, Julio C. (2003). "Validating complex construction simulation models using 3D visualization." *Systems Analysis Modeling Simulation*, 43(4), 455-467.

Khoury H.M., and Kamat V.R. (2009). "Indoor user localization for context-aware information retrieval in construction projects", *Journal of Automation in Construction*, Elsevier, 18(4), 444-457.

Law, Averill M. (2009). "How to build valid and credible simulation models", *Proceedings of the 2009 Winter Simulation Conference*, December 13-16, 2009, 24-33.

Lee, S., Han, S., and Peña-Mora, F. (2009) "Integrating Construction Operation and Context in Large-Scale Construction Using Hybrid Computer Simulation." *Journal of Computing in Civil Engineering*, ASCE, 23 (2), 75-83.

Levitt, R. E., Thomsen, J., Christiansen, T. R., Kunz, J. C., Jin, Y., and Nass, C. (1999). "Simulating project work processes and organizations: Toward a

micro-contingency theory of organizational design". *Management Science*, 45(11), 1479-1495.

Oberlender, G. D. (1997). "How do we use research to improve the engineering and construction industry?" *Final Report*, Workshop Co-sponsored by CII and NSF.

Oloufa A., Ikeda M., and Oda H. (2003) "Situational awareness of construction equipment using GPS, wireless and web technologies", *Journal of Automation in Construction*, Elsevier, 12(6), 737-748.

Peschiera, G., Taylor, J. E., and Siegel, J. A. (2010). "Response–relapse patterns of building occupant electricity consumption following exposure to personal, contextualized and occupant peer network utilization data." Energy and Buildings, 42(8), 1329-1336.

Pradhan A., Akinci B. (2012) Planning-Based Approach for Fusing Data from Multiple Sources for Construction Productivity Monitoring, ASCE Journal of Computing in Civil Engineering, 26(4), 530-540.

Proulx, G. and Richardson, J. (2002). "The Human Factor: Building Designers Often Forget How Important the Reactions of Human Occupants Are When they Specify Fire and Life Safety Systems." *Canadian Consulting Engineer*. 43(3):35-36, May issue.

Razavi S.N., Haas C.T. (2012) Reliability-Based Hybrid Data Fusion Method for Adaptive Location Estimation in Construction, Journal of Computing in Civil Engineering, 26(1), 1-10.

Rijal, H. B., Tuohy, P., Nicol, F., Humphreys, M. A., Samuel, A., and Clarke, J. (2008). "Development of an adaptive window-opening algorithm to predict the thermal comfort, energy use and overheating in buildings." *Journal of Building Performance Simulation*, 1(1), 17-30.

Robinson, S. (2004) *Simulation: the practice of model development and use*, John Wiley and Sons.

Robinson, S. and Pidd, M. (1998) "Provider and customer expectations fo successful simulation projects". *Journal of the Operational Research Society*, 49(3), 200-209.

Sampio, A. Z., Henriques, P., G., and Cruz, C., O. (2009) "Interactive models used in Civil Engineering education based on Virtual Reality technology," *Proceedings of 2nd Conference on Human System Interactions 2009*, Milan, Italy, May 21-23, 2009, 170-176.

Sergant, Robert G. (2000) "Verification, Validation, And Accreditation Of Simulation Models" *In proceedings of the 2000 Winter Simulation Conference*, J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, (eds.), Vol. 1, 41-48

Shi, J., Ren, A., and Chen, C. (2009). "Agent-based evacuation model of large public buildings under fire conditions." *Automation in Construction*, 18(3), 338-347.

Song J., Haas C.T., Caldas C.H., Ergen E., and Akinci B. (2006). "Automating the task of tracking the delivery and receipt of fabricated pipe spools in industrial projects" *Journal of Automation in Construction*, Elsevier, 15(2), 166-177.

Teizer J., Venugopal M., and Walia A. (2008) "Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and

material positioning and tracking" *Transportation Research Record* 2081(6), 56-64.

Watkins, M., Mukherjee, A., Onder, N., and Mattila, K. (2009). "Using agent-based modeling to study construction labor productivity as an emergent property of individual and crew interactions." *Journal of Construction Engineering and Management*, 135(7), 657-667.

Wall, J., and Ahmed, V. (2008) "Use of simulation game in delivering blended lifelong learning in the construction industry- Opportunities and Challenges" *Computers and Education*, Elsevier, 50, pp. 1383-1393.

Wickens, C. D. and Hollands, J. (2000). "Engineering Psychology and Human Performance." Prentice Hall

Willis, A., Gjersoe, N., Havard, C., Kerridge, J., and Kukla, R. (2004). "Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments." *Environment and Planning B Planning and Design*, 31(6), 805-828.

Xie H., and Abourizk S.M. (2011) "Integrating real time project progress input into a construction simulation model", *Proceedings of Winter Simulation Conference*, Phoenix, AZ.

Yan, W., and Kalay, Y. E. (2006). "Geometric, Cognitive and Behavioral Modeling of Environmental Users." *Design Computing and Cognition'06*, 61-79.

Yang J., Arif O., Vela P.A., Teizer J., Shi Z. (2010) Tracking multiple workers on construction sites using video cameras, Elsevier Journal of Advanced Engineering Informatics, 24(4), 428-434.

Zhang, Y., AbouRizk, S. M., Xie, H., and Moghani, E. (2012) "Design and Implementation of Loose-Coupling Visualization Components in a Distributed Construction Simulation Environment with HLA" Journal of Computing in Civil Engineering, ASCE, 25(2),248-258.

A 4D-CPM Based Graphical Scheduling System

Xing Su[1] and Hubo Cai[2]

[1]Division of Construction Engineering & Management, School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, Indiana, 47907; PH (765) 418-4331; email: xsu@purdue.edu
[2]Division of Construction Engineering & Management, School of Civil Engineering, Purdue University, 550 Stadium Mall Dr., West Lafayette, Indiana, 47907; PH (765) 494-5028; email: hubocai@purdue.edu

ABSTRACT

The emerging four dimensional modeling techniques have been proven to be useful and promising in construction planning. However, the current applications in 4D modeling still lack an efficient modeling approach, an analytical capability, and a dynamic capability. This paper proposed a 4D-CPM based graphical scheduling system aiming to address the problems. The system synergizes the creating of a conflict-free 4D model into the process of schedule development, generates and adjusts the workspace semi-automatically according to different construction methods and user options, and identifies space time conflict efficiently by analyzing the CPM network topology. It complements the traditional CPM scheduling and 3D modeling method, and could serve as an effective explanative communication tool in construction project management. The proposed system could be enhanced in three aspects: formalize a 4D topological framework, develop a 4D construction database, and enhance the collaborative and interactive capability to cater for large scale projects.

INTRODUCTION

Construction Planning and 4D Modeling Techniques. Effective planning is one of the most important aspects and directly influences the success of a construction project. It involves the choice of technology, the definition of work tasks, the estimation of the required resources and task durations, and the identification of any interactions among the different work tasks (Hendrickson, 1998).

Construction planning and control are not only temporal but also spatial. Construction site is a dynamic environment. Workspaces of construction activities, layout of temporary facilities, and material deployment continuously change on site in aspects of space and time throughout the entire lifecycle of a project. The dynamic objects on site interact with each other in a complex spatial-temporal manner. Zhou et al. (2009) stated that construction planning requires spatial information of construction elements and their workspace identification, temporal information of

schedule and duration control, and logical information for task sequence arrangement. Without a comprehensive plan that considers spatial-temporal relationships between the site objects, conflicts and improper site layout may frequently occur.

It is impractical for a planner to mentally model the dynamic process during planning. Conventional planning methods such as the Critical Path Method (CPM) consider time as the main constraint, while space limitation and space-time conflicts are minimally incorporated. The construction activity workspace execution and interference information could not be effectively represented or communicated (Mallasi, 2006). On the other hand, three-dimensional (3D) collision detection examines the spatial relationship between 3D site elements at a specific time point and therefore, it functions in a "snapshot" manner and cannot capture the temporal continuity.

Recently, research efforts are emerging in the provision of project planners and managers with computer-based advisory tools to visualize the construction plan in a 4D (3D+time) environment. Positive attitudes were received from many researchers and practitioners towards the development and application of 4D technology in construction, and the effectiveness of 4D models has been positively evaluated within different educational or industrial settings (Koo and Fischer 2000; Liston et al. 2001; and Webb et al., 2004).

Primary Research Themes in 4D Modeling. Despite the benefit and potential, 4D technology has not been accepted on a large scale in construction management (Webb et al. 2004). Some limitations of the current 4D technology are discussed such as the lack of a concept to prepare the 4D model within appropriate time and in parallel to the creation of the construction schedule (Tulke and Hanff, 2007), the lack of link to the time dimension in analysis (Mallasi, 2006), and spatial relationships and topology are not well supported by most of the 4D models (Bansal, 2011).

In summary, three issues are identified here as the primary research themes for advancing the development of 4D systems to automate construction project planning and control tasks. The first issue is **the model design and the approach of model creation**. In current practice, the 2D drawing/3D model is usually developed early in the project design phase, while the baseline construction schedules are usually generated afterward. A user typically create a 4D model by manually linking a CPM based schedule with its 3D model elements through a third party software tool (Collier and Fischer, 1996). This manual linking approach is widely applied and almost all the commercial 4D solutions adopt this technique such as Autodest Navisworks and Common Point Project 4D (Zhou et al. 2009). However, the 4D model created in this way is only for a post-plan review purpose instead of assisting the initial planning, and there is a heavy manual burden for the user due to the large number of construction elements. In addition, the separating of the 4D modeling and the scheduling activities results in very limited information and consideration in the schedule about the construction methods and production issues (Thabet et al. 1999).

To avoid these problems, many efforts have been attempted in the academic area. de Vries and Harink (2007) proposed a method for automated generation of construction planning by taking input of a 3D model and analyzing its topology. This

fully automated process completely frees users from the burden of manual linking; however, it lacks user-system interaction and the ability to provide flexible creative problem solutions. When the topology rules need to be overridden by some specific construction rules, this method may not provide appropriate results. Thabet et al. (1999) and Waly and Thabet (2003) proposed a framework "VCE" to semi-automate the scheduling process at the design phase by re-assembling a 3D CAD model using drag and drop operations. The sequence of dragging the selected 3D elements from the CAD model helps determine the temporal relationship among the activities, and the user is responsible for specifying concurrency, lags, and other relationships between activities. This interactive manual assembly approach enhances the aspect of supporting the initial planning, but the process still heavily relies on the planner's knowledge and experience during workspace identification and task arrangement.

The drawbacks of the manual assembly method bring up the second issue: **the analytical capability**. It has been identified that space-time conflict is one of the major causes of productivity loss in construction industry. However, most current 4D techniques focus on visualizing 3D graphics over time but barely support any analytical programs (Halpin 2006). A problem behind this drawback is that it lacks an efficient method to integrate the time dimension for analysis. The traditional "snapshot-scanning" styled analysis, which has to examine conflicts at every single time point, will bring up a huge computational load for a large construction project.

The second issue could be considered together with the third issue: **the dynamic capability**. Visualizing 3D objects at discrete points throughout the duration of the associated task, i.e. displaying the entire product for the whole task duration, is not representative of the construction process (Heesom and Majdjoubi 2002). The shape, location, and properties of a construction object change throughout its life cycle. For example, a cast-in-place concrete bridge pier occupies different spaces in different construction stages such as form building, concreting, and curing. As a tool for construction planning, mere product visualization provides very limited help for users to identify workspace and arrange tasks.

The three issues reveal that one of the main research trends in the area is to develop a 4D construction modeling method which has 4D dynamic and analytical capabilities and is able to be integrated into the scheduling process. The developed 4D construction modeling system should also be a decision-supporting system which is informative to help planners avoid conflicts and promote creative solutions, instead of a decision-making system which abandons human-system interactions.

A Generic 4D Construction Planning System. This paper proposed a generic 4D construction planning system which integrates space-time conflict detection and provides a GUI to assist users in 4D analyses during the initial planning phase. The system takes a 3D building product model as the input, and represents each individual activity of constructing a building element as an activity node in a CPM network. A conflict-free 4D construction model, including the space plan and schedule, could be automatically generated after user arranged and connected all the activity nodes to form a CPM network. The system provides the user with a convenient and flexible environment to create and manipulate the CPM activity nodes with workspace representation and run-time space-time conflict detection.

A 4D-CPM BASED CONSTRUCTION PLANNING SYSTEM

System overview. The 4D-CPM based construction planning system is composed of three parts: a Graphical User Interface (GUI), a Workspace Generation and Adjustment (WGA) module, and a CPM based Space-Time Conflict Detection (CPM-STCD) module. The GUI allows the user to create CPM nodes from the input 3D model to represent construction activities and connect them to form a CPM network. The WGA module generates and adjusts a workspace for each construction element semi-automatically according to different construction methods and user options. The CPM-STCD module could efficiently identify possible conflict candidates by analyzing the CPM network topology. It generates a warning message whenever a space-time conflict is detected during schedule development. Thus, the 4D model could be created simultaneously with a conflict-free schedule.

The graphical user interface (GUI). The GUI serves as an interactive platform between the user and the system. It functions as a coordinator that glues different parts and modules together. In general, it displays the 3D module as a visual clue, and provides a canvas for the user to create, arrange, and connect all the construction activities. A screenshot of the prototype is shown in Figure 1. The 3D model, which is the ground floor of a garage building, is shown in the left top corner. Each activity node in the CPM canvas is created by selecting an element of the 3D model and clicking the "Add" button. Clicking the connecting button "Con" of two activity nodes sequentially will result in a predecessor-successor relationship of the two activities. By clicking the selecting button "Sel", the detailed information of the activity will be shown in a dockable window, and the corresponding element in the 3D model will also be highlighted. The Earliest Start Time (EST) and Earliest Finish Time (EFT) are calculated and updated on run-time every time a new node is added to the CPM network. The Latest Start Time (LST) and Latest Finish Time (LFT) of activities are calculated by clicking the "Calculate" button.
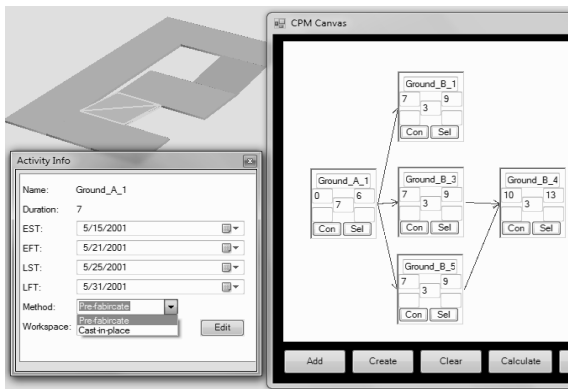

Figure 1. Prototype Screenshot.

The Workspace Generation and Adjustment (WGA) module. This module is inspired by and extended based on the 4D Work Planner Space Generator (Akinci, 2002). The geometry of a workspace is derived from not only the product geometry, but also the construction method, and it is further adjusted by its surrounding environment to ensure its correctness. Three options are provided as the basic workspace geometry generating options: buffer, rotate, and attach. Combinations of any of them can be used to represent different workspaces. The buffer option creates a shape by offsetting the original product along user specified directions. It could be used to represent workspaces of cast-in-place concreting activities and some installation activities. The rotate option rotates a shape along coordinate axes to form a shape of cylinder or sphere. The method is mainly used to help the user visualize workspaces of cranes and tower cranes. The attach option is aiming to mimic the extra workspace of labor or equipment which are required by the activity.

Figure 2 illustrates the basic examples. The light blue box in A represents a buffered workspace of the construction product; B shows the result of "buffer + attach", which buffers the construction product to create the workspace geometry then attaches it to the product or places it at a certain position specified by the user; the semi-sphere in C represents the reachable range of a crane while the cylinder in D represents that of a tower crane. The "Edit" button under the CPM canvas allows the user to fully customize the shape, in case the provided options cannot meet the requirement.
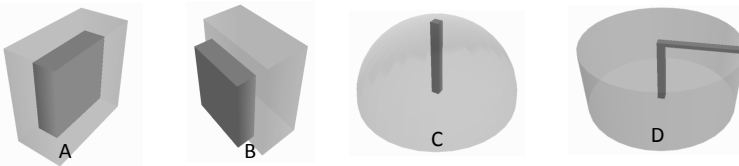

Figure 2. Sample workspace representations.

The workspace adjustment occurs when the user finds the detected space conflict is a false detection. One of the elements causing the conflict will be chosen, optionally based on the user's selection, as the compromising part. The conflict (overlapped) part will be subtracted from the geometry of the compromising element. Figure 3 illustrates an example of false detection of space conflict caused by offsetting the construction elements. The two brown parts represent two pieces of cast-in-place concrete slabs, and the two semi-transparent blue boxes represent the workspaces which are derived by offsetting the brown parts. If the right slab is concreted first, the left slab concreting can only start after the right slab has been cured. In this case it will be detected that the *workspace* of the left slab has space conflict with the right slab's *product* shape, however, this is a false detection. If the user decides to adjust it, the overlapped geometry will be subtracted from the workspace geometry automatically. If both of the conflict geometries are *workspaces* (not *product*), the user will have to determine which one is the compromising part.
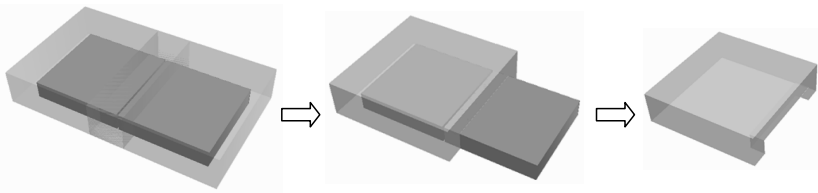
Figure 3. An example of workspace adjustment.

The CPM Based Space Time Conflict Detection (CPM-STCD) Module.
The space time conflict in this research specifically means the workspace interference when two activities have both spatial and temporal overlaps. In the 4D CPM network, every node is treated as a 4D object carrying both space and time attributes. In other words, every node represents a construction activity which takes a certain length of time and occupies a certain workspace. Space-time conflict detection conducted directly on the 4D objects avoids the effort of scanning all the snapshots along time at every single time point, which could improve the efficiency and make 4D optimization possible. In this module, the detection is triggered every time the user connects two activity nodes.

The time dimension resides within the logical connection and sequence of a CPM network, thus the topology of a CPM network becomes the key to decoding each activity's time attribute. A CPM network may have a complex structure; however, the topology of its nodes always consists of some basic relationships. Some related terms are defined here as follows:

1. Predecessor and Successor: If A points to B, A is B's predecessor, and B is A's successor;
2. Indirect-Predecessor and Indirect-Successor: If B is A's successor, all nodes after B, except for A's successors, are A's indirect-successors; and all nodes before A, except for B's predecessors, are B's indirect-predecessors;
3. Pre-Neighbor and Post-Neighbor: If A and B all point to C, A and B are pre-neighbors of C; If C points to A and B, A and B are post-neighbors of C;
4. Neighbor branch: If A and B are post-neighbors, all A's successors and indirect successors, except for B's successors and indirect successors, are in B's neighbor branch of A.

As displayed in Figure 4, B is A's successor and D's predecessor. A's indirect-successors include D, E, F, and G. C is excluded because it is also A's successor. F's indirect-predecessor include A and B. B and C are post-neighbors of A; and E and F are pre-neighbors of G. C has one neighbor branch which is comprised of D and F. Sometimes two nodes could have more than 1 relationship. For example, B and C are post-neighbors of A; in the mean while C is B's successor.
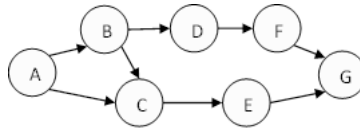
Figure 4. A sample CPM network.

The space-time conflict detection includes three steps: time-conflict detection based on CPM topology, time-conflict detection based on EST and EFT, and space-conflict detection based on geometry. An empty candidate list is created before the three steps. The first step is to detect possible time-conflict nodes based on CPM topology. The logical dependency of a CPM network determines that a node will never have any time-conflict with its successors and indirect successors. However, a node is time-conflict possible with its post-neighbors and nodes in its neighbor branches. These pairs of nodes are added to the candidate list.

A node's neighbor branch may go through the CPM network to the very end, and it is inefficient to list every node in its neighbor branch as time-conflict pairs. The candidate list could be shortened by the second step, time-conflict detection based on EST and EFT, which filters out pairs of nodes with time gaps. Since the EST and EFT of each node are calculated once the node is added to the network, the time gap between two nodes could be determined immediately by checking whether the late node's EST is later than the early node's EFT. Once a node in the neighbor branch of the original node is found to have a time gap with the original node, the node in the neighbor branch and all of its successors and indirect successors have no time conflict with the original node.

The last step is to perform space-conflict detection on all the pairs of nodes in the candidate list. Each node is associated with two 3D geometries representing the workspace and the product shape respectively. Collision detection is conducted on the workspace and product shapes of the two nodes. The CPM-STCD module ensures a conflict free 4D model. After the user creates the CPM network, all information required by a schedule or 4D simulation could be found in the activities' attributes such as ID, duration, workspace geometry, product geometry, EST, EFT, LST, and LFT. Some of the temporal and semantic attributes could be written as outputs in desired formats by looping through all the activities.

CONCLUSION AND DISCUSSION

The proposed system is expected to benefit the construction domain in four aspects: a) the system integrates the traditional CPM scheduling and 3D modeling method to form a 4D CPM-based framework, which laid a foundation for a generic construction 4D planning system, b) the graphical scheduling interface can synergize the creating of a 4D model into the process of schedule development, c) a workspace generation and adjustment method is developed to flexibly represent the real world space usage, and d) space-time conflicts could be efficiently detected on run-time. From the practical perspective, the CPM-based system, which is compatible with the widely used construction scheduling method, can facilitate the implementation in industries.

The prototype of the system is currently under development and being enhanced towards a generic 4D construction planning system. The future research directions include formalizing a 4D topological framework for construction, developing a 4D construction database system that supports topological, geometrical, and semantic query and analysis, and enhancing the collaborative and interactive capability to cater for large scale projects.

REFERENCES

Bansal V. K. (2011). "Use of gis and topology in the identification and resolution of space conflicts." *J. Comput. Civ. Eng.,* 22(4), 233-242.

Collier, E. and Fischer, M. (1996). "Visual-based scheduling: 4D modeling on the San Mateo County Health Center." *Proceedings of the 3rd ASCE Congress on Computing in Civil Engineering,* Anaheim, CA, 800-805.

de Varis, B., and Harink, J. M. J. (2007). "Generation of a construction planning from a 3D CAD model." *Autom. Constr*., Vol. 16(1), 13-18.

Halpin, D. W. (2006). "Construction Management." Chapter 7, Project Scheduling, John Wiley & Sons, Inc. 101-125.

Heesom, D. and Mahdjoubi, L. (2002). "Visualization development for the virtual construction site." Dessertation, School of Engineering and the Built Environment, University of Wolverhampton.

Hendrickson, C. (1998). "Project management for construction – fundamental concepts for owners, engineers, architects and builders." available at http://www.ce.cmu.edu/pmbook/.

Koo, B., and Fischer, M. (2000). "Feasibility study of 4D CAD in commercial construction." *J. Constr. Eng. Manage.,* 126(4), 251-260.

Liston, K., Fischer, M. and Winograd, T. (2001). "Focused sharing of information for multi-disciplinary decision making by project teams," *Electronic Journal of Information Technology in Construction*, 6, 69–81.

Mallasi, Z. (2006). "Dynamic quantification and analysis of the construction workspace congestion utilizing 4D visualization." *Autom. Constr*. 15, 640-655.

Thabet, W., Y., Wakefield. R., R., and Walf, A., F. (1999). "Virtual construction for Automated schedule generation." *Durability of building materials and components 8*. Institute for research in construction, Ottawa ON, K1A 0R6, Canada, pp. 2939-2944.

Tulke J., Hanff J. (2007). "4D construction sequence Planning -- new process and data model." *CIB 24th W78 Conference*. Maribor, Slovenia, 79-84.

Waly, A. F., and Thabet. W. Y. (2003). "A virtual construction environment for preconstruction planning." *Autom. Constr*. 12, 139-154.

Webb, R. M., Smartwood, J., and Haupt, T. C. (2004). "The potential of 4D CAD as a tool for construction management*." J. Constr. Res.,* 5(1), 43-60.

Zhou, W., Heesom, D., Georgakis, P., Nwagboso, C., and Feng, A. (2009). "An interactive approach to collaborative 4D construction planning." *ITcon*, Vol. 14, 30-47

A semiotic analysis of the meaning encoding, decoding, and interpretation process during the development and use of building information systems

T. Hartmann, A.M. ASCE[1]

[1] VISICO Center, Department of Construction Management and Engineering, University of Twente, P.O. Box 217, Enschede, 7500AE, The Netherlands; PH +31(53) 4893376; email: t.hartmann@utwente.nl

ABSTRACT

This paper presents a unified semiotic theory of computer based communication for the build environment. The developed theory understands (computer) sign based communication as a triadic process involving real world objects, the signs computers use to represent them, and interpretants that represent the understanding that users were able to produce about the built environment by interpreting these signs. The triadic character of the semiotic theory allows us to understand the cascading process of developing and using BI systems during which actors create computer based signs to represent objects in the built environment (semantic representation), how others interpret signs, and how, based, on these interpretations these others can then encode the interpreted meaning using new signs. Next to developing the above outlined semiotic meaning decoding, interpretation, and encoding process framework, the paper also shows by example that this semiotic framework allows understanding the process of how signs allow for the communication of meaning about the built environment during the development and use of BI systems.

INTRODUCTION

The last years have seen increasing focus on the research and development of Building Information (BI) based computer systems. These systems promise to support and streamline the communication between building designers, increase their decision making capabilities, and support the communication with non-experts. Despite this inherent benefits of BI systems to support communication, research has so far mainly focused on the social and organizational process of how ready-made BI systems can support communication (Tomo and Cerovsek, 2011). Few efforts exist that explore the underlying semiotic mechanisms and processes of how BI systems are, can be, or should be used to convey design meaning and knowledge. Understanding such semiotic mechanisms better, however, is of utmost importance because at their core systems that are built around information models representing buildings are semiotic tools that encapsulate knowledge about buildings and should support the communication of this knowledge. Next to organizational and managerial process analyses, in depth semiotic analyses of BI systems are, therefore, required to provide meaningful insight about how to best develop and use BI systems that truly support

the communication between professionals working in the architectural, design, and construction industry.

To complement prior work in this area with the goal to support such analyses (Mutis 2012, Hartmann 2012, Hartmann forthcoming a, Hartmann forthcoming b), this paper introduces a theoretical framework to describe the meaning decoding, interpretation, and encoding process during the development and use of BI systems by drawing upon computer semiotic theories (Andersen, 1997; Andersen et al., 1994; de Souza et al., 2001). The framework focuses on the different levels at which BI software systems encapsulate and communicate knowledge: The engineering representation level, the BI schema level, the BI model level, and the BI system user interface level. The framework also accounts for the main actors that are involved in the communication of knowledge: The software development and design professionals, the AEC engineering professionals, and the laypersons that need to be involved in construction projects. To show the power of semiotic framework, the paper illustrates the process model using an exemplary BIM development and use process from our ongoing research efforts in the area of modeling inner city sewer construction projects.

The paper is structured as follows: It starts with a brief introduction of computer semiotic theory and a discussion of how this general theory can be used to understand BI systems. Next, the paper introduces the semiotic framework to analyze the meaning decoding, interpretation, and encoding processes during the development and use of BI systems. Subsequently, the paper illustrates the applicability of the framework to understand the development of BI systems better. The paper concludes with a discussion of the illustrations.

COMPUTER SEMIOTICS

Semiotics is the study of signs and their use to convey social meaning (Andersen, 1997) where signs are defined as ``something which stands to somebody for something in some respect or capacity'' (Pierce, 1955). This definition of a sign is of a triadic nature, in that it is assumed that a sign is composed of three parts: the *object* represented, the *representamen* that represents the *object*, and the *interpretant* which is the mental representation of the *representamen* by a person (Andersen, 1995). This triadic nature of the theory unifies three important parts of how a sign can convey meaning: (a) the syntax level of the sign, e.g., its shape, form, and appearance, (b) the semiotic of the sign, e.g., how it uses its shape, form, and appearance to represent the object, and (c) the pragmatic function of the sign, e.g., how it helps somebody to make meaning about the object by producing a *representamen* with the help of the sign (Hasle, 1994).

Semiotics has been applied in many different theoretical fields and strands, but important in the context of this paper is the strand of computer semiotics (Andersen, 1997). Computers semiotics is concerned with how computers use signs to encode and transmit meaningful signals from human senders to human receivers (Andersen, 1997). Computer semiotics considers software not as processing devices, but as a knowledge encoding media. The knowledge that the software encodes cascades through all abstraction levels of computer systems. Low level programming languages encode knowledge about how to store and execute programs using the

provided computer hardware. Semantic object models used to program the software using programming languages encode the knowledge of the programmers about the domain that the software should support. The graphical user interface of the software encodes the intentions of the programmers of how to best use the system and also allows software users to encode their knowledge to communicate with other users. With this view, computer semiotics allows to move away from the traditional understanding that computers are merely batch processing systems towards considering computers as semiotic systems that are part of the social interaction between humans (Piotrowski, 1994).

Against this backdrop, the value of computer semiotics is not in developing universal principles for designing computer systems better, but in allowing for a better understanding of computer related processes and products. Computer semiotics aims at providing system designers and users of computer systems with new perceptions that help them to analyze specific systems within specific, but varying, contexts. To this end, independently from the abstraction levels described above, the field of computer semiotics tries to use sign theory to understand, describe, and inform the different types of knowledge encoding interaction involved during the design of a BI system (de Souza et al., 2001):

- *Engineer - system developer* interaction during the development of the BI system,
- *Engineer - system interaction* during the design of a domain specific BI model, and
- *Engineer - engineer or engineer layperson* interaction during the design of domain overarching BI models.

To support such interaction within the Architectural, Engineering, and Construction domain meaningfully with BI systems, semiotic signs have to be developed and interpreted on the different semiotic levels of a BI system. Figure 1 tries to theoretically capture and describe this process. Traditionally, engineers interpret "real objects" and encode this knowledge in traditional engineering representations, such as two-dimensional drawings, sketches, or specifications. During the development of BI systems, developers have to in a first step understand these engineering interpretations in relation with physical objects within the built environment to identify the requirements of the system. Once interpreted developers can encode these objects again representing them as formal BI model schema (see Zhang and Issa, forthcoming; Demian and Balatsoukas, 2012; Nepal et al., forthcoming; Anil et al., 2012, for some recent examples of such schema). This interpretation is then symbolically implemented in a BI software system (for how to semiotically analyze such symbolic implementation the interested reader can refer to Hartmann, forthcoming a). This symbolic implementation of data is then interpreted by engineers who can instantiate a specific model instance of the symbolized building object to support a specific design effort. This final representation of the model instance within the system can then be interpreted by computer software to provide computer based representations that other engineers and laypersons can then decode and interpret again. In the end, ideally, this final interpretation by the engineer or lay person allows

them to understand the meaning about the physical building object encoded during the previous cascades of the overall meaning encoding and interpretation process.
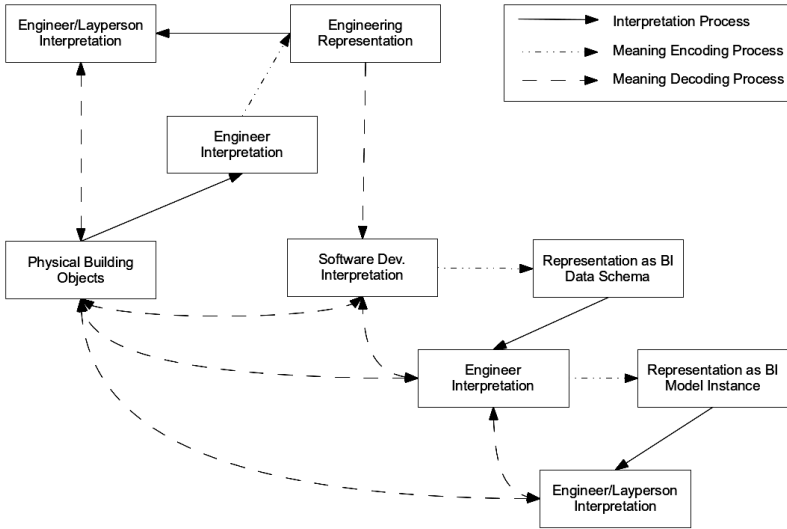

Figure 1: The meaning encoding, decoding, and interpretation process during the development of a BI system.

To provide first illustration of the mechanism that this process framework sketches, the next section illustrates an application of the framework by analyzing ongoing research efforts to develop a BI system for the planning of inner city sewer construction projects at the authors research group.

USING THE FRAMEWORK TO ANALYSE THE DEVELOPMENT OF A BI SYSTEM: THE CASE OF A PLANNING SUPPORT SYSTEM FOR INNER CITY SEWER CONSTRUCTION PROJECTS

I use the above described framework to analyze the transfer of meaning within BI systems by describing the efforts of an ongoing research project. This research effort aims at developing a BI system with the purpose to support decision making activities during planning inner city sewer construction projects. Figure 2 shows a representation of the different semiotic stages following the format of Figure 1.

At the logical start of every BI system development trajectory stand the objects of the physical world, in this case, the physical objects existing on a typical sewer construction site, such as pipes, cables, excavation pits, or lanterns (Figure 2-a). Following traditional practice, engineers on inner city sewer construction projects represented those real object that were of relevance during the design and construction of the sewer line using engineering representations, similar to the here

illustrated part of a two dimensional construction drawing (Figure 2-b). These drawings itself already represent interesting sign systems, illustrated in Figure 2-b for example, by the textual annotations or notes that were sketched on the drawings to symbolically represent characteristics of real world objects that cannot be represented well using two dimensional sign systems.
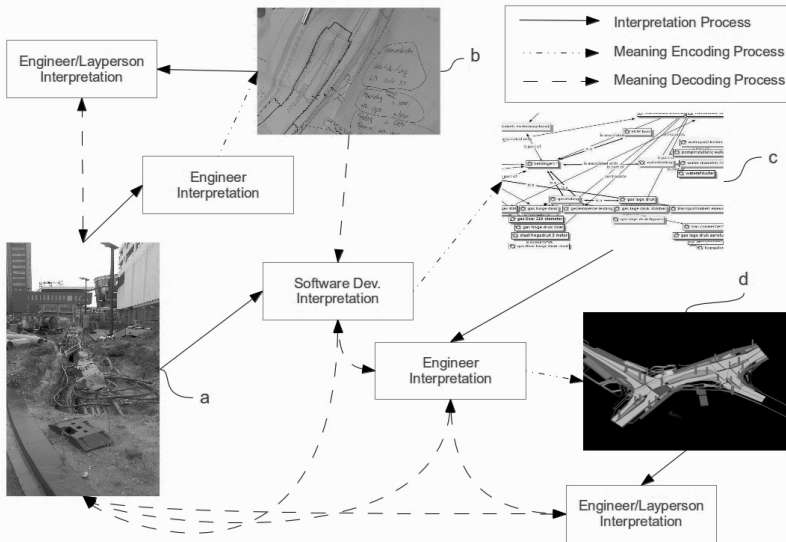

Figure 2: Illustrative meaning decoding, interpretation, and encoding process during the development and use of a BI system to support inner city sewer projects. Call outs represent real world objects (a) and several semiotic sign systems used to encode meaning. These semiotic sign system cascade meaning during the process and have to be decoded and interpreted (b – two dimensional drawing as an example of a traditional engineering representation; c – formal ontology to represent inner city sewer projects in general; d – computer generated 3D representation of a BI model for a specific inner city sewer construction project).

Using these traditional engineering representations as a starting point, the research team interpreted both the real world objects (Figure 2-a) and the existing engineering representation of the objects (Figure 2-b) to understand the for the decision making activities of the engineers on inner sewer projects most important objects with their characteristics and relations. The research team then encoded these objects and their characteristics in a BI data schema (Figure 2-c). Semiotically this schema represents a generalization of concepts within a specific domain – an ontology. Ontologies are descriptions that are detached from real world context representations and try to describe objects, their characteristics, and relations with each other (Uschold and Gruninger, 1996). Once the research team had develop such an ontology through an

in depth interpretation of engineering representations and their interpretations of their own observations of the existing objects on inner city sewer construction projects, the ontology was then coded as an object oriented data model. For engineers to make use of this object oriented data models, they in a first step need to interpret the generalized schema and use it to instantiate a concrete model for a specific project. This interpretation and instantiation is usually done in BI model authoring applications (Eastman et al., 2008). On the example project, the research team provided such a BI authoring tool on the basis of an existing CAD software. Using this tool, it was now possible to represent specific projects in a BI model instance by using interpretations of the ontology. In the next cascade of the semiotic meaning interpretation and encoding process, computer applications can then interpret these BI model instances and again represent these automatic representations symbolically (Figure 2-d). In a final step, these symbolic computer interpretations can now again be interpreted by engineers or laypersons to understand specific of the real world objects.

DISCUSSION
If nothing more, the above semiotic analysis of a development and use process of a BI system reveals that the described development process cascaded through a number of different meaning interpretation and encoding steps. With this insight, I show that the here presented semiotic framework can already provide guidance during BI development and use processes and makes the required steps during these processes explicit. This, in turn, will be helpful to manage these processes on research and development projects that intend to implement new BI systems.
Understanding the meaning interpretation and encoding mechanisms during the development of BI systems explicitly is also important because such understanding can draw specific attention to possible problems that might occur during the development and use of the system. Developers and users, for example, should be aware that within each of the above sketched cascades there is the risk that failures can occur during the required encoding, interpretation or decoding steps. Following semiotic theory such failures occur due to misunderstandings of semiotic systems and can be caused by two main reasons (Andersen, 1997):

- The signs used by the computer system conform to the principles of the semiotic system as envisioned by the developers, but upstream users of the signs that might be developers, engineers, or laypersons fail to find a relation to what is to be signified by the semiotic system.
- The signs used in the computer system do not conform to the principles of the semiotic system as it is intended during the encoding of the semiotic meaning system.

Besides these practical implications for management BI system development and use efforts directly, from a more theoretical viewpoint, the introduced framework and its underlying assumption contributes with assuming that BI systems are, at their core, meaning encoding systems. This new understanding of BI system can shed new light on our understanding of BI systems as a whole. Most BI systems are developed to

support communication between system engineers and laypersons to allow for better engineering decision making processes. Semiotic theories, such as the here presented theory, can help to understand better how the existing BI tools already support communication and how to develop new BI systems that do so even better. Hereby, semiotic theories cannot help to provide hard-coded implementation guidelines. Rather they provide meaningful frameworks to support developers more generally by allowing them to gain new perceptions on the underlying mechanisms of the communication processes their software should support. The here presented work presents such a framework that complements other semiotic frameworks developed recently (Mutis 2012, Hartmann 2012, Hartmann forthcoming a, Hartmann forthcoming b). Additionally, these semiotic theories help users to better understand how BI system already work or can and should work. Such understanding will then allow them to more meaningfully use BI systems for the decision making and communication tasks.

CONCLUSION

This paper introduced a framework that describes the different meaning interpretation and encoding mechanisms during the development and use of BI systems. The framework draws on computer semiotic theory and intends to provide new perceptions about the underlying mechanisms of how BI systems can be developed and used to meaningfully support communication and decision making tasks on construction projects. The paper also provided an illustrative application of the framework by analyzing the research and development efforts to develop a BI system to support inner city sewer design and construction activities. The example illustrates how the framework allows for a better understanding of the overall development and use process of BI systems and how an analysis of such development and use processes with the framework can draw attention to specific pitfalls during the encoding, decoding and interpretation of meaning within BI systems.

Concluding, the framework presented is a further stepping stone that allows the civil engineering community to re-conceptualize BI systems from information processing devices to meaning encoding devices. In the end, such semiotic theories will allow developers and users to better understand the underlying mechanism of how BI systems can convey meaning. Such an understanding will, in turn, lead to better possibilities to develop and use BI systems to truly support complex multi-disciplinary engineering decisions.

REFERENCES

Andersen, P. (1995). "The force dynamics of interactive systems: Toward a computer semiotics." Semiotica, 103(1-2), 5–46.

Andersen, P. (1997). A theory of computer semiotics: semiotic approaches to construction and assessment of computer systems. Cambridge series on human-computer interaction. Cambridge University Press.

Andersen, P., Holmqvist, B., and Jensen, J. (1994). The Computer as Medium. Learning in Doing. Cambridge University Press.

Anil, E., Unal, G., and Kurc, O. (2012). "Information requirements for design and detailing of reinforced concrete frames in multiuser environments." Journal of Computing in Civil Engineering, 26(4), 465–477.

de Souza, C., Barbosa, S., and da Silva, S. (2001). "Semiotic engineering principles for evaluating end-user programming environments." Interacting with Computers, 13(4), 467 – 495.

de Souza, C., Barbosa, S., and Prates, R. (2001). "A semiotic engineering approach to user interface design." Knowledge-Based Systems, 14(8), 461–465.

Demian, P. and Balatsoukas, P. (2012). "Information retrieval from civil engineering repositories: Importance of context and granularity." Journal of Computing in Civil Engineering, 26(6), 727–740.

Eastman, C., Teicholz, P., Sacks, R., and Liston, K. (2008). BIM Handbook – A guide to Building Information Mpdeling for Owners, Managers, Designers, Engineers, and Contractors. Wiley & Sons, Inc., Hoboken, New Jersey

Hartmann, T. (2012) "A Semiotic Analysis of Building Information Model Systems" in Proceedings of the International Conference on Computing in Civil Engineering: Computing in Civil Engineering (eds. Issa, R. R. and Flood, O.): 381-388.

Hartmann, T. (submitted a). "A semiotic analysis of user interfaces of building information systems." Submitted to the Journal of Computing in Civil Engineering.

Hartmann, T. (submitted b). "A semiotic framework to understand how signs in construction process simulations convey information". Submitted to Advanced Engineering Informatics.

Hasle, P. (1994). "Logic grammer and the triadic sign relation." The Computer as Medium, Cambridge University Press, 104–127.

Mutis, I. (2012). "i-Con - Geometric Topologies for Semantic Interpretation of Building Components Based on a Semiotic Framework" in Proceedings of the International Conference on Computing in Civil Engineering: Computing in Civil Engineering (eds. Issa, R. R. and Flood, O.): 17-24.

Nepal, M., Staub-French, S., Pottinger, R., and Zhang, J. (forthcoming). "Ontology-based feature modeling for construction information extraction from a building information model." Journal of Computing in Civil Engineering.

Pierce, C. (1955). "Logic as semiotic: The theory of signs." The philosophical writings of Pierce, 98–119.

Piotrowski, D. (1994). "Structuralism, computation and cognition. the contribution glossematics." The Computer as Medium, Cambridge University Press, 68–91.

Tomo and Cerovsek (2011). "A review and outlook for a building information model (BIM): A multi-standpoint framework for technological development." Advanced Engineering Informatics, 25(2), 224 – 244.

Uschold, M. and Gruninger, M. (1996). "Ontologies: Principles, Methods, and Applications." Knolwedge Engineering Review, 11(2).

Zhang, L. and Issa, R. (forthcoming). "Ontology based partial building information model extraction." Journal of Computing in Civil Engineering.

A Pedagogical Benchmark Experiment for Application of Multidisciplinary Design Optimization in Early Stage Building Design

Shih-Hsin Eve Lin[1] and David Jason Gerber[2]

[1]Ph.D. Candidate, School of Architecture, University of Southern California, Watt Hall, Suite 204, Los Angeles, CA 90089-0291; email: shihhsil@usc.edu
[2] Assistant Professor, School of Architecture, University of Southern California, Watt Hall, Suite 204, Los Angeles, CA 90089-0291; email: dgerber@usc.edu

ABSTRACT

This research is built upon a previously established early stage multidisciplinary design optimization (MDO) framework, entitled Evolutionary Energy Performance Feedback for Design (EEPFD), and proceeds with observing the impact of EEPFD on the early stages of design by conducting a pedagogical benchmark experiment. This experiment has two main observational interests. The first objective is to observe discrepancies between the human versus automated decision making processes and the resulting performance of the solution space from each process. The second objective is to understand students' ability to translate their design intent into a parametric model, as this is a crucial component in the implementation of EEPFD. By comparing EEPFD and the benchmark pedagogical process, this paper provides an initial assessment of the potential of EEPFD to reduce latency in decision making and to find superior performing design solutions compared to the benchmark process. At the completion of this experiment it was observed that EEPFD was able to deliver superior performing design solution spaces, but that students encountered difficulties in the translation of their design intent into a functioning parametric model.

INTRODUCTION + RESEARCH OBJECTIVE

Buildings consume nearly half (49%) of all energy used by the United States. Building Operations alone account for 43.5% of U.S. energy consumed today while construction and building materials account for an additional 5.5% (Architecture 2030 2011). However, the overall performance of the building is greatly impacted by design decisions made in the early stages of the design process, when design professionals are often unable to explore design alternatives and their impact on energy consumption (Schlueter and Thesseling 2009, Flager et al. 2009).

Research precedents have demonstrated the potential of adopting MDO to provide a performance feedback loop for supporting early design stage decision making (Flager et al. 2009, Welle, Haymaker, and Rogers 2011). However, precedents exploring MDO in the AEC field have typically employed simplified geometry (Flager et al. 2009, Welle, Haymaker, and Rogers 2011) while precedents involving more complex geometry have limited themselves to single domain

optimization (Yi and Malkawi 2009). Where the energy performance domain has been included for optimization the relationship between design form and energy performance has been largely excluded. Furthermore, the application of these precedents' subject of interest to the overall design process remains largely unexplored.

In response to this gap in existing research a MDO design framework was developed that could incorporate both conceptual energy analysis and exploration of complex geometry for the purpose of providing early stage design performance feedback. Subsequently, the established MDO design framework can then be applied to the overall design process where its impact can then be observed.

The established MDO design framework, EEPFD, stands for "Evolutionary Energy Performance Feedback for Design," utilizes the prototype tool H.D.S. Beagle which enables the coupling of parametric design with multi-objective optimization (Gerber and Lin 2012a, Gerber et al. 2012, Gerber and Lin 2012b). EEPFD specifically enables the exploration of varying degrees of geometric complexity with energy simulation feedback. Also provided are spatial programing compliance and financial performances for consideration in performance tradeoff studies. EEPFD has demonstrated the ability to reduce design cycle latency, automate design exploration, analysis and the evaluation process, and provide improving performance design alternatives. In addition, the established MDO framework has begun testing in pedagogical and practical design process settings (Gerber and Lin 2012a, Gerber et al. 2012, Gerber and Lin 2012b).

This paper presents the impact of EEPFD on the design process through a benchmark pedagogical experiment. The pedagogical experiment is conducted within a design computational tool course environment with students who are in the process of learning both simulation tools and their application to design. The interest of this experimental set is to observe any measurable effects of the introduction of EEPFD sans the element of automation enabled by the Beagle. The second primary observational interest for this experiment set is with regards to the students' ability to translate their design intent into a parametric model for further exploration.

INTRODUCTION OF EEPFD + H.D.S. BEAGLE

EEPFD, a MDO based design framework, was developed in parallel to H.D.S. Beagle and can be considered a proposed means of implementation of the concepts driving the Beagle in direct application to the design process. In order to realize EEPFD, selection of the platforms by which to implement the GA-based multi-objective optimization (MOO) algorithm was needed. To this end a prototype tool (H.D.S. Beagle) was developed as a plugin for Autodesk® Revit® which integrated Autodesk® Green Building Studio® (GBS) and Microsoft® Excel® to generate the desired automation and optimization routine.

Autodesk® Revit® is a building information modeling platform with parametric capabilities enabling designers to define their geometry while providing a series of parameters that impact the development of varying geometric configurations. This platform also serves as an insertion point for the energy settings necessary for a conceptual energy analysis through Autodesk® Green Building Studio® (GBS). GBS is a web-based energy analysis service that serves as the energy simulation engine for

the prototype. Microsoft® Excel® 2010 provides not only a means of containing the financial parameters and formula, but also as a user interface proxy in which designers can set up design parameter ranges of interest, constraints, spatial program parameters, and the spatial programing compliance formula. The three objective functions can be formulaically expressed as follows:

S_{obj} = Max. SPC   where
E_{obj} = Min. EUI S_{obj} = *Spatial Programming Compliance Objective Function*
F_{obj} = Max. NPV E_{obj} = *Energy Performance Objective Function*
                        F_{obj} = *Financial Performance Objective Function*
                        SPC = Spatial Programming Compliance Socre
                        EUI = Energy Use Intensity
                        NPV = Net Present Value

A more detailed description of the adopted method which drives EEPFD and the automated engine of H.D.S. Beagle can be found in previously published research (Gerber and Lin 2012a, Gerber et al. 2012, Gerber and Lin 2012b).

EXPERIMENT DESIGN & METHOD

This experiment was conducted by the authors through the course entitled Arch507: Theories of Computer Technology during the Spring Semester of 2012 at the School of Architecture of University of Southern California (SoA USC). The course is currently offered to both undergraduate and graduate students of architectural design, building science, structural engineering and construction management disciplines. A total 27 students participated in the experiment. The student group was composed of 17 Master of Architecture candidates, 8 Master of Building Science candidates, 1 undergraduate student pursuing a Minor in Architecture, and 1 Master of Building Science graduate.

The pedagogical experiment is designed as a benchmark case for the evaluation of EEPFD. To this end the simulation process used by EEPFD requires adjustment so as to be suitable for manual use within the pedagogical experiment. Figure 1 illustrates both the simulation process used by EEPFD and the adjusted process designed for the pedagogical context. The significant simulation process difference is predicated on the fact that the pedagogical experiment is run in a classroom setting where students are not granted access in real time, or in parallel, to use of the Beagle and in particular its cloud based architecture and automation. Therefore, the last three steps need to be performed manually by the participants as opposed to automated through the GA-based multi-objective optimization process provided though the Beagle.

After determination of the simulation process, the research organizes the experiment into three major activities: 1) course lecture and hands-on practice by students; 2) assignment; and 3) obtaining data from students for analysis and cross comparison. All activities are designed to utilize the same platforms as EEPFD so students' exploration results can be considered comparable to results obtained through EEPFD. As a result, the contents of the lecture and assignment are divided into four major portions: 1) Parametric Modeling: The method by which to create a

parametric model using Revit's conceptual massing environment. 2) Objective Function Calculation: The method by which to conduct Conceptual Energy Analysis, identify relevant information within the Revit model, and input this relevant information into the Excel objective function calculator so as to calculate the three objective functions: EUI, SPC & NPV. 3) Design Exploration: The process through which students can manipulate their parametric model and explore their design alternatives. 4) Design Evaluation: The method of evaluating design alternatives according to the Pareto Rank Evaluation method (Gerber et al. 2012).
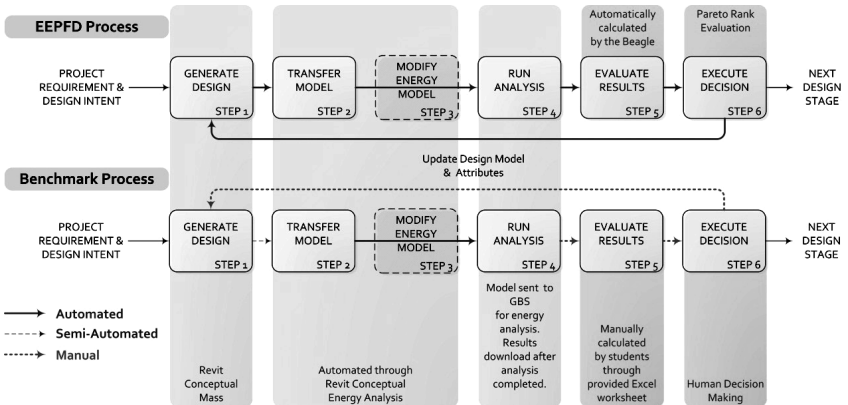
Figure 1. Simulation process outline for EEPFD and the pedagogical experiment
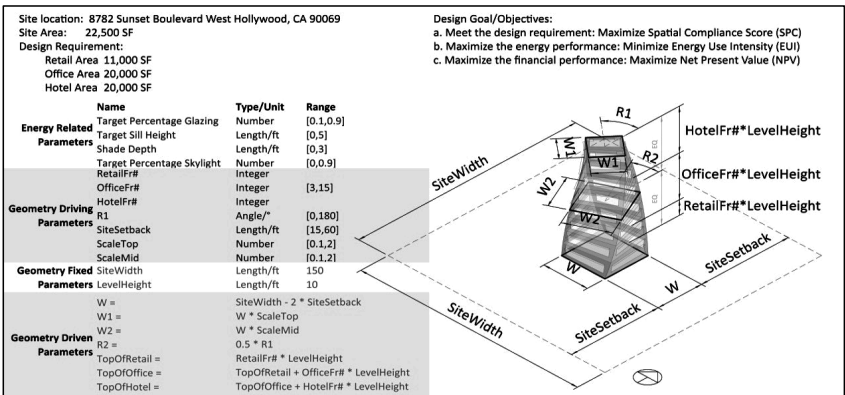
Figure 2. The provided design requirements, parametric model and ranges for the pedagogical benchmark experiment for Part A.

After the lecture students are asked to complete an assignment following the adjusted simulation process as instructed. The assignment is divided in two primary parts: Part A and Part B. Both parts are given identical design requirements to ensure comparability of results to those generated by EEPFD. However, in Part A of the

assignment students are asked to explore a given parametric design, as illustrated in Figure 2, with the objective of identifying a higher performing design alternative within the given parametric ranges. In Part B students are asked to generate their own design intent and then translate this intent into a parametric model. At this point students are instructed to utilize a total of five form driving parameter. FloorNumber and SiteSetback are required and enable the use of three fully customized form driving parameters unique to each student's design. This allows for the observation of the translation process of design intent into a parametric model as this is a critical component of EEPFD. For both Part A and Part B the types of data recorded are described in Table 1.

Table 1. Summary of the recorded data for the overall pedagogical experiment

	Recorded Data	Data Type
Background Questionnaire	1. Education Background 2. Experience using Revit 3. Experience using other parametric software	Enumeration Number (Time) Y/N Number (Time) Enumeration
Part A & Part B	1. Explored parametric values for each iteration 2. The three objective function scores of explored design alternatives 3. Time spent to obtain and calculate each objective function 4. Time spent to obtain the final design 5. Design exploration process	Number Number (EUI, SPC, NPV) Number (Time) Number (Time) Image
Additional Recorded Data in Part B	1. Three design intents prior to modeling their own parametric model 2. Design parameterization process map 3. Parametric model design 4. Created parameters 5. Initial value and variation range of each parameter 6. Ability of parameters to represent desired design intent	Description Image Revit File Enumeration numbers Y/N Percentage

EXPERIMENT RESULTS & OBSERVATIONS

The following is a selected summary of the results extracted from the recorded data as relevant to the previously described objective of this paper. During Part A students recorded an average exploration time of 3 hours, 5 generated iterations, and approximately 16.4 minutes per iteration. Therefore, for qualitative comparison purposes, the Beagle was given the same time period in which to generate results for the same design problem. However, in the interest of exploring the capability of the Beagle when provided extended resources, the Beagle was also given a 7 hour runtime and asked to reach 6 generations in separate runs.

For comparison purposes there are two result sets of interest. The first is in comparing the range of student generated solution space (SGSS) with the Beagle generated solution space (BGSS) within the designated 3 hour time period. The second set is to further include BGSS after 7 hours and 6 generations runs into comparison. Table 1 provides a summary of the performance ranges of the solution pools generated by these result sets.

When comparing the solution pools generated by students and the Beagle under the same time constrains, the Beagle was able to provide a solution pool with a 26.8% increase in the measured NPV and a 13.7% reduction in the calculated EUI.

However, the student solution pool was able to provide a 22.2% increased SPC score. When ranked according to all three objective performance 36.7% of the design alternatives generated by the Beagle were designated as Pareto optimal solutions while only 26.9% of the students generated design alternatives received this designation. Further improvement in the performance of the solution pool was observed when the Beagle was provided an extended period of time, as reflecting of potential availability through increased computational affordances.

Table 2. Performance comparison of solution spaces generated by the pedagogical experiment and the Beagle after completed runtimes of 3 hours, 7 hours, and 6 generations.

			SOLUTION SPACE RANGE		
		SGSS	BGSS		
			3 HRS	7HRS	6th Gen
NPV	MIN.	-48	-41	-41	-41
(Million$)	MAX.	341	432	584	835
Initial: -41	Improved	382	473	626	875
EUI	MIN.	67	59	59	56
(kBtu/sqft/yr)	MAX.	292	233	233	233
Initial: 174	Improved	107.0	115.1	115.1	117.8
SPC	MIN.	-151	-134	-266	-404
	MAX.	99	81	83	88
Initial: 10	Improved	90	72	73	78
Pareto (%)(Students/Beagle)	3 OBJ		26.9/36.7	22.4/40.0	19.4/37.2

In Pedagogical Experiment Part B data from 25 students was collected. Time spent for this part of the experiment by students averaged 5.5 hours with a range of 1 to 13 hours. Recorded time spent includes both initial model set up time and design exploration time through a minimum of four design iterations. At the completion of the experiment students were asked to gauge their success in translating their design intent into a parametric model. Overall, 86% of the students responded that they felt they were able to successfully translate their design intent into a parametric model. However, this rate of success was not reflected in the overall evaluation of all received models.

Table 3 provides a summary of the overall assessment of the parametric models as received by students at the end of the experiment. Parameter Accuracy pertains to the quality and capability of all student provided parameters by observing the correlation of each parameter and the actual impact on the design geometry. Model Robustness is used to determine whether or not a student's design model is able to maintain its integrity throughout the design exploration process. This assessment is made through evaluating the parametric ranges and rules as provided by each student for their individual design model. Driving Parameter Compliance evaluates whether students followed the given instructions regarding their form driving parameters to ensure a consistent comparable quantity of explored parameters among the students' designs models. Accuracy of Calculated Data was used to validate student generated results by confirming both proper model set for simulation purposes and calculations based on simulated results. Finally, a final evaluation was made in order to assess the overall quality of the student provided models.

One key observation is with regards to the Accuracy of Calculated Data results which demonstrated a significant issue regarding human error in the calculation process. In addition, only 40% of the students were able to generate an accurate parametric model with which to generate accurate objective scores for their designs. Furthermore it can be observed that the issue regarding how to ensure Model Robustness requires direct response.

Table 3. Evaluation summary of students' parametric model performance

Evaluated Category	Parameter Accuracy	Model Robustness		Driving Parameter Compliance	Accuracy of Calculated Data		Final Evaluation
		Rule	Range		Setup	Calculation	
Scale	(Poor, Acceptable, Good)	(Y/N)	(Y/N)	(Y/N)	(Y/N)	(Y/N)	(A-D)
Summary	P: 36% A: 24% G: 40%	N: 44% Y: 56%	N: 100% Y: 0%	N: 52% Y: 48%	N: 44% Y: 56%	N: 52% Y: 48%	A: 0% B: 56% C: 24% D: 20%

CONCLUSION

During the comparative studies of the experiment, EEPFD demonstrated the ability to generate superior results under the same time constraints as human users, and further improved results within the designated performance objectives when given extended time. Given that time typically dominates early design exploration it can be extrapolated that the reduction in computation time necessary to generate desired results would further acclimate the framework to the early stage design process. In addition it was observed that EEPFD using the Beagle is able to negate issues regarding human based error as observed in the manual calculation of the objective scores. It can be extrapolated then that with increasingly complex design problems there is a greater potential for human error and therefore an increased benefit to the automated process available through EEPFD. This would enable more informed early stage design decision making for even more complicated design projects, which is a subject for future studies.

However, these advantages are dependent on the prerequisite component of a functioning parametric model suitable for exploration through EEPFD. One initial observation regarding the translation of design intent into a parametric model as described by the students was the disparity between perceived success by students and actual functioning results compatible with EEPFD and H.D.S. Beagle. While students considered themselves successful in translating their design intent into a parametric model the quality of the resulting parametric model and its ability to reflect the described design intent were typically found lacking in critical components. Increased experience regarding the composure of a functioning parametric model may be able to close this disparity gap and is in need of further research.

One element that defies parametric translation is that of aesthetic preference which differs widely between individuals. As a result resources will typically be expended on optimizing only potential design solutions that satisfy this first element. As the exploration process of EEPFD possesses no aesthetic preference equally it possesses no aesthetic prejudice. While it may spend time analyzing solutions that will ultimately be dismissed by the designer, equally it analyzes solutions potentially

overlooked by the designer. The result is a broader based design solution pool with over all improved multi-objective performance levels to enable more informed design decision making inclusive of a more expansive simulated aesthetic and formal range.

ACKNOWLEDGEMENT

REFERENCES

Architecture 2030. 2011. *Energy - Buildings consume more energy than any other sector* 2011 [cited 18 April 18 2012 2011]. Available from http://architecture2030.org/the_problem/problem_energy.

Flager, Forest, Benjamin Welle, Prasun Bansal, Grant Soremekun, and John Haymaker. 2009. "Multidisciplinary process integration and design optimization of a classroom building." *Information Technology in Construction* no. 14 (38):595-612.

Gerber, David Jason, and Shih-Hsin E. Lin. 2012a. Designing-in performance through parameterisation, automation, and evolutionary algorithms: 'H.D.S. BEAGLE 1.0'. Paper read at CAADRIA 2012, 25-28 April 2012, at Chennai, India.

Gerber, David Jason, and Shih-Hsin Eve Lin. 2012b. Synthesizing design performance: An evolutionary approach to multidisciplinary design search. Paper read at ACADIA 2012 - Synthetic Digital Ecologies 18-21 October 2012, at San Francisco, California, USA.

Gerber, David Jason, Shih-Hsin Eve Lin, Bei Penny Pan, and Aslihan Senel Solmaz. 2012. Design optioneering: Multi-disciplinary design optimization through parameterization, domain integration and automation of a genetic algorithm. Paper read at SimAUD 2012, 26-30 March 2012, at Orlando, Florida, USA.

Schlueter, Arno, and Frank Thesseling. 2009. "Building information model based energy/exergy performance assessment in early design stages." *Automation in Construction* no. 18 (2):153-163. doi: 10.1016/j.autcon.2008.07.003.

Welle, Benjamin, John Haymaker, and Zack Rogers. 2011. "ThermalOpt: A methodology for automated BIM-based multidisciplinary thermal simulation for use in optimization environments." *Building Simulation* no. 4 (4):293-313. doi: 10.1007/s12273-011-0052-5.

Yi, Yun Kyu, and Ali M. Malkawi. 2009. "Optimizing building form for energy performance based on hierarchical geometry relation." *Automation in Construction* no. 18 (6):825-833. doi: 10.1016/j.autcon.2009.03.006.

Computing Geometric and Mass Properties of Culturally Important Statues for Rigid Body Rocking

C.E. Wittich[1] and T.C. Hutchinson[1]

[1]Dept. of Structural Engineering, University of California, San Diego, 9500 Gilman Drive, Mail Code 0085, La Jolla, CA 92093; email: {cwittich, tara}@ucsd.edu

ABSTRACT

The seismic response of statues is motivated by observations from recent earthquakes combined with their high cultural significance. Typically unrestrained with a high coefficient of static friction at its base, a statue can be analyzed as a rigid body which will undergo rocking during seismic excitation. In this paper, a methodology is presented to obtain the critical geometric and mass properties required for numerical rigid body rocking analyses. Two nonintrusive approaches to capture the statue geometry are presented: light detection and ranging (LiDAR) and structure-from-motion (SfM), which generate a three-dimensional point cloud. A Poisson reconstruction algorithm intrinsically filters each point cloud and creates a fully enclosed triangulated surface mesh. Integrating over the surface mesh allows for computation of the required geometric and mass properties for numerical rocking analysis to simulate the seismic response of a statue.

INTRODUCTION

The analysis of cultural heritage artifacts, statues in particular, has been a largely neglected area of structural and earthquake engineering with advances only in the past decade or so (Nigbor 1989). Yet, their high cultural significance combined with observations from recent earthquakes gives impetus to their study (e.g. Rosetto 2012) as statues are susceptible to toppling during earthquake excitation. This is due to the statues' physical attributes which generally include tall aspect ratios, construction from of a single piece of marble, and unrestrained placements on a stone pedestal with a high coefficient of static friction (Wittich et al., 2012).

Monolithic bodies, such as statues, can be analyzed in the context of rigid body dynamics, namely rigid body rocking. Nigbor (1989) conducted a survey of statues within a museum collection and assessed their vulnerability through rigid body rocking analysis using envelope shapes and estimated mass properties. Although studies have taken advantage of approximate ranges of statue geometries for understanding the general rocking response of statues (Kounadis 2012), advances in computational power support development of more detailed computer models of a statue. For example, in the late 1990s the Digital Michelangelo Project (Levoy 2009) successfully created three-dimensional digital reconstructions of statues with a high

degree of accuracy. The statues were scanned using a laser triangulation range finder with a motorized arm for controlling placement producing a point cloud consisting of millions of points each consisting of nodal coordinates, as well as color and intensity values. In this study, millimeter point cloud resolution was obtained, with minimal occlusion, and post-processing techniques such as filtering and surface meshing were utilized (Levoy 1999). Resampling has resulted in sub-millimeter resolution (1/4mm) for the famous David statue. Due to the availability of laser scanners in recent years, higher accuracy models for statue analysis are more commonplace. Berto et al. (2012) presented a general preservation analysis of a set of statues in Florence, Italy, which included laser scans of statues, meshing, and solid element finite element analysis. Given the complex nature of statue geometry, a careful methodology must be applied to maintain accuracy in these digital renderings. Herein, this paper presents a clearly outlined methodology for obtaining geometric data of complex statues and gleaning relevant information from this data. To this end, two methods for obtaining point clouds of statues are presented along with their transformation into triangulated surface meshes from which geometric and mass properties are calculated.

RIGID BODY ROCKING

Culturally important marble statues subject to earthquake excitation can be expected to undergo a rocking-dominated response. This is due to their large aspect ratios (i.e. $3 < H/B < 5$, where H = height to the center of mass and B = base dimension) and high coefficients of static friction ($\mu_s \approx 0.75$) between the base of the statue and the top of their pedestal. The equation of motion for a two dimensional rigid block was initially derived by Housner (1963) and later re-derived and analyzed by others (Ishiyama1982; Shenton 1996). Figure 1 includes an image of a culturally important statue, the *Florentine Pieta* by Michelangelo, overlaid with an equivalent two-dimensional rigid block including the geometric parameters relevant to the equation of motion: $\ddot{\theta} = -\frac{mRg}{I_{rp}}\left[\text{sgn}(\theta)\sin(\alpha - |\theta|) + \frac{\ddot{u}_g}{g}\cos(\alpha - |\theta|)\right]$, where θ is the angle of rotation, m is the mass of the block, g is the acceleration due to gravity, R is the distance from the center of mass to the point of rocking (rp), I_{rp} is the mass moment of inertia about the rocking point, sgn is the signum function, α is the critical angle, and $\ddot{u}_g$ is the acceleration of the ground in the horizontal direction. In reality, statues are three-dimensional eccentric rigid bodies subject to three-dimensional ground excitation. However, two-dimensional examples for the equation of motion and illustration of parameters are provided in this section for simplification and to convey the types of geometric and mass properties necessary for rigid body rocking
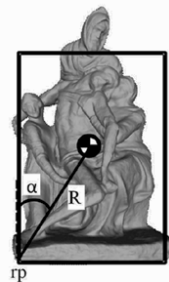

Figure 1. Two-dimensional rigid block with relevant geometric parameters with underlying image of representative statue.

analysis. Eccentricity can be accounted for by unique geometric parameters for a particular rocking point. The set of equations of motion for a three-dimensional rocking rigid body contains the same parameters, where a derivation can be found in Zulli et al. (2012). Irrespective of whether eccentricity and higher dimensions are accounted for, the critical parameters are location of the center of mass, dimensions of the footprint, and moment of inertia about all possible rocking points.

DATA ACQUISITION

The complex geometry of statues combined with the highly nonlinear equations describing rocking motion requires the acquisition of accurate three-dimensional digital renderings. Acquisition of point cloud data is presented herein to be obtained by one of two methods, namely, light detection and ranging (LiDAR) or structure-from-motion (SfM). LiDAR is a very accurate measurement system that uses laser technology to collect a series of points representing the physical structure in its view with respect to location (x,y,z), color (r,g,b), and intensity. Millimeter and sub-millimeter resolution can be obtained with current LiDAR technology. A more cost-effective yet more photorealistic approach is SfM reconstruction. SfM consists of identifying matching features within a set of images and triangulating to form a point cloud. This method is used in conjunction with scaling to obtain a geometric representation; however, the resolution is not uniform and noise in the point data can be significant which becomes evident in the following sub-sections.

LiDAR. Light detection and ranging (LiDAR) systems comprise a broad spectrum of laser-emitting devices used for range measurement, or more typically known as laser scanning. Three-dimensional laser scanning is the most accurate method for recording geometry commercially available today. A typical scanner transmits and receives laser pulses or waves to record distance as a first step; and as a second step, an embedded digital camera photographs the scene so that color and texture may be applied to the depth map. The scanner used for acquisition of the surveyed statues in this work is the FARO Focus 120 3D, a phase-based scanner. Phase-based LiDAR scanners constantly emit and receive continuous wave laser beams while recording the phase shift of the wave after reflectance in order to determine distance to the object in line of sight. The Focus 3D has a range of 120m and the potential for sub-millimeter resolution given the spot size of 3.8mm, the step size of 0.009°, and the ability for close range scanning of the statues (< 1m in the horizontal plane). For an object scanned at a maximum distance of 2m, the noise is estimated to be 0.12mm uncompressed and the error to be ±0.4mm for a marble surface (FARO 2011). Ten retroreflective spheres were positioned for redundancy and alignment of multiple scans. Scans were aligned using the spheres in the corresponding FARO Scene software. A representative aligned statue point cloud is shown in Figure 2.

SfM. Structure-from-Motion (SfM) is three-dimensional scene reconstruction using images from cameras in the field, reconstructed using feature-tracking algorithms on the computer as a post-processing step. The technique is based upon

the theory of passive stereo, which dictates that given two projections of the same point in space, its location relative to the two rays can be determined through triangulation. SfM applies this concept to multiple points or features within a given set of photographs to triangulate the location of the cameras. Feature or point tracking in the SfM algorithms searches for uniqueness of a pixel or a group of pixels from the surrounding pixels in terms of color, edges, corners, shadows, and high contrast. The feature points combined with the relative position to the cameras create an unscaled three-dimensional reconstruction of an object.

For the statues in this study, three cameras were used for digital capture for SfM. The cameras were: (1) Canon EOS 7D with a EF-S 28-135mm f/3.5-5.6 IS lens, (2) Canon EOS 5D Mark II with a EF 24-105mm f/4L IS lens, and (3) Nikon D7000 with a 18-105mm f/3.5-5.6 VR lens. It should be noted that only one high-quality camera and lens would be sufficient for this survey; however, three were used due to multiple photographers and equipment availability. The approximate cost of the equipment required for this survey is $1000 US in 2011 compared to approximately $25,000 US in 2011 for the FARO system for LiDAR. Both techniques require approximately 1 hour of field time for data acquisition. For each statue, an attempt was made to capture the object digitally from above and from below for optimal coverage and linking correspondences between images.

The statues surveyed were reconstructed from images using VisualSfM with corresponding dense reconstruction algorithm (Furukawa and Ponce 2010; Wu 2011). Image distortion is accounted for as an initial step in the dense reconstruction phase. The resultant point clouds were scaled according to pedestal dimensions obtained in a field survey. A representative reconstruction is presented in Figure 2 for comparison with that of LiDAR. Significant noise around the surface of the statue can be observed with the SfM results; however, the texture and color are more realistic, which may not be essential for global rocking analysis under earthquake scenarios, however, it is important for cultural heritage documentation. The total number of vertices and point densities that describe the point clouds obtained by SfM and LiDAR are included in Table 1. SfM is able to produce a point cloud approximately 10-15% as dense as that of LiDAR; however, this value will be clarified in the following section regarding filtering and surface meshing.

SURFACE MESH

Prior to calculation of any geometric or mass properties, the point clouds must be transformed into enclosed surface meshes typically comprised of triangles. Due to the inherent noise associated with the acquisition methods, as well as the extremely complex surface topology of the statues, a Poisson surface reconstruction algorithm was applied to each point cloud. Kazhdan et al. (2006) presented this type of surface reconstruction in the context of sculpture meshing from noisy real-world data sets. The comparisons given indicate that Poisson reconstruction of statues from LiDAR data is more efficient and produces higher detail surfaces than the other common reconstruction algorithms such as Volumetric Range Image Processing (VRIP) and Fourier transform based approaches. This type of algorithm simultaneously creates a watertight, enclosed, triangulated surface mesh and filters the point cloud for surface
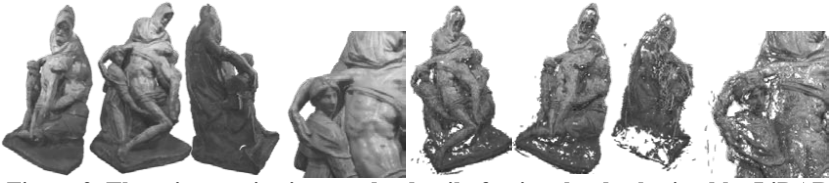
Figure 2. Three isometric views and a detail of point clouds obtained by LiDAR (left) and SfM (right) for the *Florentine Pieta* by Michelangelo.

Table 1. Comparison of point density of raw point clouds and triangulated meshes for SfM and LiDAR for two surveyed statues.

	Statue 1 F. Pieta by Michelangelo			Statue 2 Zuccone by Donatello		
	# Points	Surface Area [m^2]	Avg. Point Density [pts/m^2]	# Points	Surface Area [m^2]	Avg. Point Density [pts/m^2]
Raw Point Clouds						
SfM	558895	9.76*	66220	438050	3.74*	141764
LiDAR (1mm)	3399339	9.76*	348293	3766116	3.74*	1006983
Triangulated Meshes						
SfM	21387	8.44	2534	9371	3.09	3033
LiDAR (1mm)	735561	9.76	75365	818626	3.74	218884

NOTE: (1mm) = a decimated resolution of 1mm for the underlying point cloud.
Surface area is taken as the value obtained using LiDAR to compare average point density.

noise, making it the ideal option for LiDAR or SfM data of complex statues. Furthermore, no common point cloud filtering techniques are available for such detailed surfaces with sharp gradient changes, as is present for statues. This intrinsic filtering step is evident by the count of points and point density associated with the raw point clouds compared to that of the surface meshes associated with both LiDAR and SfM. It is also worth noting that after filtering, the point densities associated with SfM are reduced to 1-3% that of LiDAR compared to that of the pre-filtered point clouds at 10-15% (Table 1) emphasizing the level of noise associated with SfM. The time associated with filtering and meshing SfM is approximately 5 hours per statue compared to the resultant point clouds of LiDAR which require only about 1 hour.

The final step in the processing of the point clouds was to resample the mesh and subsequently the underlying point cloud in order to recover a resolution via a clustering decimation algorithm. A three-dimensional grid of specified size surrounds the mesh and collapses the vertices in accordance with the size of the grid. Both Poisson reconstruction and clustering decimation were implemented in Meshlab (Visual Computing Lab 2011). Meshes and corresponding point clouds resulting from LiDAR scanning were able to be decimated to a uniform resolution of 1 mm neglecting areas of occlusion which were estimated. Figure 3 compares the resultant meshes associated with LiDAR and SfM. These results emphasize the fine detail captured with LiDAR, whereas only the general shape is captured with SfM. A detailed comparison associated with mass properties resultant from varying degrees of meshing for both LiDAR and SfM is presented in the following section.

MASS PROPERTIES

Given a fully enclosed triangulated mesh as previously introduced, the final step to calculate geometric and mass properties relevant to rigid body rocking is conducted. The moment of inertia and centroid are computed by integration over the mass. However with a surface mesh, this direct integration is not possible, but with an assumption of constant density (reasonable for a marble statue) the mass integrals are easily transformed into a volume integral. With use of the divergence theorem, volume integrals
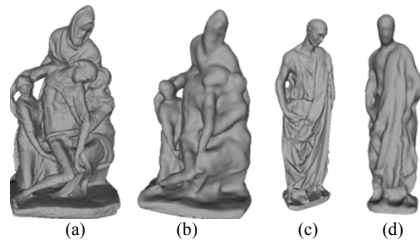

(a)   (b)   (c)   (d)

Figure 3. Poisson reconstructed triangulated meshes of *F. Pieta* by Michelangelo (a,b) and *Zuccone* by Donatello (c,d) obtained by LiDAR (a,c) and SfM (b,d).

can be reduced to surface integrals; however, their implementation is difficult and time intensive. Mirtich (1996) and Eberly (2009) presented algorithms for fast and accurate implementation of these surface integrals specific for triangulated surface meshes for mass, centroid, and moment of inertia. For the statue meshes with an assumed constant density, three properties are yielded by the integration: mass, location of the centroid, and moment of inertia about the centroid. However for rigid body rocking analysis, the moment of inertia needs to be translated such that it is with respect to the rocking points, and the distance from the centroid to the rocking point and the critical angle (α) must be determined. Each of these properties requires locating the rocking points. The first step to define the rocking points (edges) is to determine the edge vertices. This is accomplished by identifying the triangles with the lowest height, and subsequently identifying the triangles which have vertices at the extrema. For the two-dimensional rectangular footprint case, the rocking point is determined as the average of the identified edge points in the direction normal to the plane of rocking. For three-dimensional cases and non-rectangular statue footprints, both rocking edges and points must be determined. As the number of corners on typical statues will be low (4-6), the most efficient method begins with estimated locations of the rocking points. Linear regression is then used to fit the rocking edges. The locations of the rocking points are then identified by iterations for the best fit to the rocking edges. This was implemented in MATLAB in conjunction with the integration for the mass, centroid, and inertia tensor.

Table 2 includes values of computed geometric and mass properties for one representative statue obtained by envelope shape (bounding box – where the bounding planes are defined by the furthest most point in a given direction), SfM, and LiDAR at three decimated resolutions (1mm, 3mm, and 10mm). For the two coarser LiDAR meshes, the parameters are calculated within 5% that of the finest resolution. Furthermore, SfM calculated within 10% of the finest resolution of LiDAR. Envelope shapes produced deviations of greater than 100% and are biased toward overestimation of mass properties and underestimation of slenderness. Mass property analysis of other culturally important statues surveyed in this study demonstrated

similar deviations when using bounding boxes compared to both LiDAR and SfM (Wittich et al. 2012). Ground acceleration required for initiation of rocking is directly correlated with the location of center of mass and slenderness; and, a deviation in the these properties yields a similar deviation in the acceleration required for rocking initiation. A 10% deviation in peak ground acceleration does not produce an appreciable deviation in the percentage of exceedance of a particular seismic event; and, therefore, this level is considered acceptable for SfM. Ongoing studies by the authors aim to quantify the effect of variations in geometry for rocking and determine whether this level of deviation is acceptable.

Table 2. Comparison of height of center of mass (COM_z), mass moment of inertia (I_{zz}), mass (m), and critical angle (α_{xz} and α_{yz}) calculated by envelope shape (bounding box), SfM, and LiDAR of *Florentine Pieta* by Michelangelo.

		Envelope	SfM	LiDAR (10mm)	LiDAR (3mm)	LiDAR (1mm)
COM_z	[m]	1.1854	0.9675	0.9676	0.9211	0.9237
	[% diff]	28.3%	4.7%	4.8%	-0.3%	
I_{zz}	[kg·m^2]	887	343	387	387	388
	[% diff]	128.6%	-11.6%	-0.3%	-0.3%	
m	[kg]	5756	2801	2901	2902	2903
	[% diff]	98.3%	-3.5%	-0.1%	0.0%	
α_{xz}	[deg]	18.6	20.5	18.3	18.8	18.7
	[% diff]	-0.5%	9.6%	-2.1%	0.5%	
α_{yz}	[deg]	25.0	32.2	31.0	32.0	31.8
	[% diff]	-21.4%	1.3%	-2.5%	0.6%	

NOTE: 1. % diff is calculated with respect to properties of LiDAR (1mm); 2. Envelope dimensions are: 1.1m x0.8m x2.4m; 3. Density of marble is assumed at 2760 kg/m^3; 4. Point densities for SfM and 1mm are approx. 2500 and 75000

CONCLUSIONS

Due to the nonlinearity of the rocking equation of motion, accurate methods for determining the geometric and mass properties are critical. In this work, a methodology for determining geometric and mass properties of complex structures is presented within the context of historical statues. It is recognized that such structures have highly complex geometries, and in the majority of situations are tall and slender and therefore susceptible to damage due to overturning under seismic loading. LiDAR and SfM can be used to obtain point clouds combined with a Poisson reconstruction algorithm to create an enclosed triangulated surface mesh. Transforming the mass integrals into surface integrals allows for calculation of the necessary statue parameters. Structure-from-Motion reconstruction was shown to reasonably capture the geometric and mass properties of the statues when compared to that of LiDAR; whereas, the use of envelope shapes and estimation with no resultant point cloud cannot produce the properties with any reliable degree of accuracy. The results of this work can be used to guide future acquisitions of three-dimensional digital reconstructions of complex structures.

ACKNOWLEDGEMENTS

REFERENCES

Berto, L., Favaretto, T., Saetta, A., Antonelli, F., and Lazzarini, L. (2012). "Assess seismic vulnerability of art objects: The 'Galleria dei Prigioni' sculptures at the Accademia Gallery in Florence." *J. Cult. Herit.*, 13(1), 7-21.

Eberly, D. (2009). "Polyhedral mass properties (revisited)." http://www.geometric tools. com/Documentation/Documentation.html (Jan. 8, 2013)

FARO. (2011). *FARO Laser Scanner Focus 3D: Features, Benefits, & Technical Specifications*. FARO Technologies, Inc., Lake Mary, FL.

Furukawa, Y. and Ponce, J. (2010). "Accurate, Dense, and Robust Multi-View Stereopsis". *IEEE T. Pattern Anal.* 32 (8), 1362-1376.

Housner, G.W. (1963). "The behavior of inverted pendulum structures during earthquakes". *Bull. Seismol. Soc. Am.*, 53(2), 403-417.

Ishiyama, Y. (1982). "Motions of rigid bodies and criteria for overturning by earthquake excitation". *Earthquake Eng Struc*, 10(5), 636-650.

Kounadis, A.N. (2012). "Parametric study in rocking instability of a rigid block under harmonic ground pulse." *Soil Dyn. Earthq. Eng.*, 45(2), 125-143.

Levoy, M. et al. (1999) "The Digital Michelangelo Project". *Proceedings of ACM SIGGRAPH 2000*, New Orleans, LA.

Levoy, M. (2009). *Digital Michelangelo Project*. <graphics.stanford.edu/mich>

Mirtich, B. (1996). "Fast and accurate computation of polyhedral mass properties." *J. Graph. Tools*, 1(2), 31-50.

Nigbor, R.L. (1989). *Analytical/experimental evaluation of seismic mitigation measures for art objects*. PhD Thesis, University of Southern California.

Rosetto, T. et al. (2012). *The 20th May 2012 Emilia Romagna Earthquake*. EPICentre Field Observation Report EPI-FO-200512, London, UK.

Shenton III, H. W. (1996). "Criteria for initiation of slide, rock, and slide-rock rigid-body modes". *J. Eng. Mech.-ASCE*, ASCE, 122(7), 690-693.

Visual Computing Lab. (2011). *Meshlab*. http://meshlab.sourceforge.net/

Wittich, C.E., Hutchinson, T.C., Wood, R.L., Kuester, F., and Seracini, M. (2012). *Survey and Characterization of Culturally Important Statues in Florence, Italy*. SSRP 2012/10, University of California San Diego, La Jolla, CA.

Wu, C. (2011). *VisualSFM*. http://homes.cs.washington.edu/~ccwu/vsfm/.

Zulli, D., Contento, A., and DiEgidio, A. (2012). "3D model of rigid block with a rectangular base subject to pulse-type excitation." *Int. J. Nonlinear Mech.*, 47(6), 679-687.

Real-Time Bricks Counting for Construction Progress Monitoring

L. Hui[1] and I. Brilakis[2]

[1]Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK; PH (44) 0774 6140845; email: lh469@cam.ac.uk

[2]Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK; PH (44) 1223 332718; email: ib340@cam.ac.uk

ABSTRACT

When constructing brick façades, the prevalent method of measuring progress is through manual site surveys. These surveys are tedious and time consuming. They are also approximate, as counting the number of bricks in place to compare against those ordered is a very laborious task compared to its end value. This paper presents a novel and automatic method for counting bricks in place to automate the brick site survey during façade construction. This method works by using images or video frames taken at a construction site. Color thresholds are selected based on the color of the bricks, and are used on the video frames to distinguish bricks in the image. The edge map of the frames is generated using a Laplacian of Gaussian filter. The edges are compared with the known brick properties, including their rectangular shape, their expected arrangement and their size. Test results demonstrate that this method is capable of counting the number of bricks on a brick façade with acceptable error.

Keywords: Brick count; Progress monitoring; Brick detection; Site management

INTRODUCTION

Construction progress monitoring is necessary for estimating the progress of a project from time to time to support decision making and alter the successor activities in an effective way. Traditional progress monitoring is mainly based on the daily, weekly and monthly progress reports provided by the site manager and workers (United-States-Census-Bureau 2012). Manual inspection and site survey are needed to support the reports. However, these methods are highly dependent on the personal experience of the inspector. It is only an approximation of progress. According to a survey(Assaf and Al-Hejji 2006), inefficient planning and scheduling by contractor is one of the top five causes of delay stated by the owners and consultants, and the late procurement of materials is one of the top ten causes of delay by contractors. According to the Vermont Government Website(Vermont-Government-Website 2003), 20% of the trash from the state each year is construction trash, including

90,000 tons of construction and demolition waste which ends up in landfills. Current project progress monitoring practices are based on manual estimation and therefore the control of progress and material procurement are based on the approximations. A new progress monitoring method with precise results is needed to improve the existing situation.

In this paper, we use the case of brick façade construction, and present an automated brick counting method for installed bricks to provide an accurate numerical estimation of progress. We propose to use visual data taken from the site to count the number of bricks on a brick façade. The results show that the number of bricks can be counted using their color, shape and size.

BACKGROUND

State of Practice

Traditional brick facade progress monitoring is based on manual inspection at the construction site. Progress is mainly estimated based on the completion of activities, signposts or milestones set in the schedule plan. The contractor provides progress reports daily, weekly and/or monthly to facilitate project schedule management and report to the owners. The report includes the actual jobs that are completed and an estimation of percentage of completeness (United-States-Census-Bureau 2012). The estimation in the reports is performed visually and approximately. The contractors need to inspect the site and estimate according to their experience. The resulting reports provide an approximation of progress and written descriptions of the work.

Manual site surveys do not provide a precise estimate of progress. The site manager can only have an approximated view of the progress, and the decisions on procurement orders, handling the waste produced and controlling the project schedule are made based on this approximation. Therefore, the site manager is not able to have a precise plan for the schedule change and material order.

Visual Based Detection

He Junji et al. (2010) proposed a method to detect the size of firebricks using canny edge detection to extract the contour shape of bricks from the firebrick images. The edge points extracted are then linked, segmented and fitted with lines. The length and width of a given brick can be calculated by the characteristics of the camera. This paper shows that the size of a brick can be detected by the image of the brick. However, the method is focusing on the manufacturing images which have only one single brick. The method cannot be used on a brick façade with a large quantity of bricks.

Schmitt et al. (2000) proposed a method to control brick quality by detecting the defects, included surface indentations or protrusions, cracks, and surface of a brick in a uniform pink color. The result shows that it is possible to extract the edge of bricks from the image, and perform detection based on the color of bricks.

However, this detection is also only focusing on images of single brick, and is not able to handle brick facades.

The research question that this paper will attempt to answer is: how to detect and count the number of bricks on a brick façade? The objective of this paper is to test whether a novel method that we propose can reasonably detect and count the number of bricks in a brick façade.

The scope of this work is to automate the detection of bricks and count their number on a brick façade for as-built progress monitoring. We assume the following visual properties of bricks:

- Rectangular shape
- Minor variations due to defects and light conditions are acceptable
- The boundary is enclosed
- The bricks are similar in size

Most of the bricks used in brick façade are of rectangular shape, similar in size or identical, and have an enclosed boundary. Therefore the shape is assumed to be the standard shape size and arrangement of the bricks. The light conditions will vary in the outdoor sites due to the weather and time of the day, so the method should be able to tolerate a variance in the light conditions, assuming that the images are taken during day-time.

METHODOLOGY

We propose the following method to detect and count the number of bricks using the visual properties of bricks, a Laplacian of Gaussian filter and color thresholding. The number of bricks in a façade is proportional to the construction progress. Therefore counting the number of bricks can give a better estimate of the construction progress. The detailed methodology is presented in figure 1.
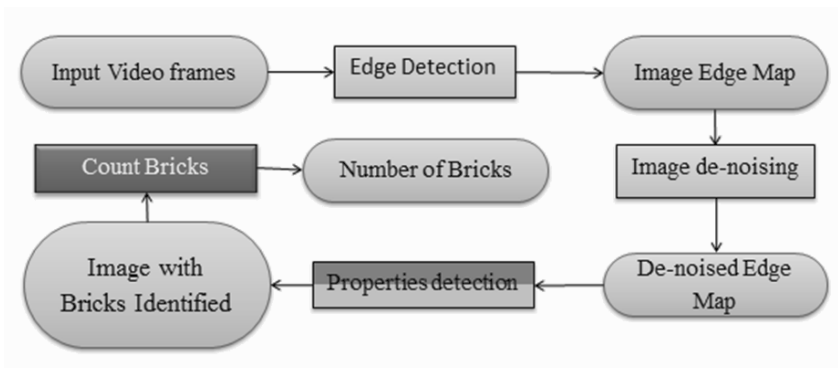

Figure 1 Work flow of the brick detection algorithm

Before counting the number of bricks, they should be detected and located.

The image or video frames are input into the system for detection. The color threshold value will be chosen based on the user indication of a few bricks in the

image. The threshold restricts the color into a range of RGB values in the chosen bricks, and produces a binary image where the color within the RGB range is in white and the rest is in black.

Using this binary image, we generate the edge map of the elements inside the image, and use the edge map as a contour shape indicator. The edge map is shown in figure 2.

We consider the assumptions of rectangular shape, fully enclosed contour edge, and that bricks on a façade are of similar size to detect the bricks. The rectangular shape property is checked by the four right angles in one continuous contour in the image. The area
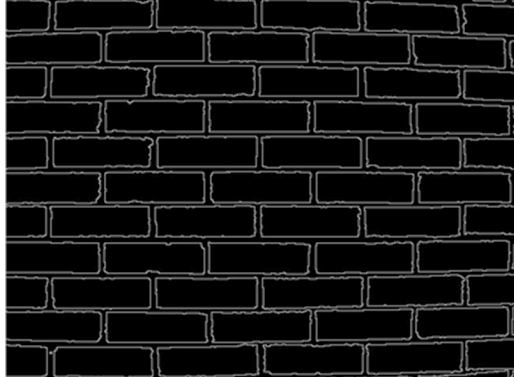

Figure 2 Edge Map of a Façade

is checked to be within ±50% from the mean area of the rectangular contours. The duplicated counts are deleted if the two contours overlap. The number of bricks can be counted from the detected bricks; the detection and counting results are highlighted and printed on the display. The methodology is limited to the bricks with a distinguishable size to show an enclosed rectangular boundary in the given pixels. The relative area of a single brick to the image would not affect the performance of the methodology.

IMPLEMENTATION AND RESULTS

The methodology presented was developed and tested on brick walls with red bricks. The images were taken under natural daylight conditions.

For edge extraction, we choose to test Laplacian of Gaussian and Canny edge detection filters. The edge maps of an image can be seen in figure 3:

Original Figure   Laplacian of Gaussian   Canny edge detection

Figure 3 Camera rotation

The two edge detection operators can both detect brick edges successfully, therefore both of them can be used as the filter. Considering the assumptions of

rectangular shape and area requirement, Laplacian of Gaussian produces a more uniform and true to area output than Canny edge detection. Therefore Laplacian of Gaussian is chosen to be our edge extraction filter. The de-noising methodology chosen is morphological opening.

We used two stages to test the methodology. The proposed methodology was coded and implemented using C#. The testing images used are of the size of 640 by 480 pixels, taken from a brick façade structure in the campus of the Georgia Institute of Technology.

The method performance is measured by determining the precision and the recall of the results. The precision is an indicator of the correctness of the detection, and the recall is an indicator of the completeness of the detection. After running the experiment, the True Positive (TP), False Positive (FP) and False Negative (FN) detections are determined, where TP is the number of correctly detected bricks, FP is the number of falsely detected bricks and FN is the number of non-detected bricks. The precision is calculated by $\frac{TP}{TP+FP}$ and the recall by $\frac{TP}{TP+FN}$. Bricks with 50% or greater of their surface included in the image but not detected were assumed false positive. The implementation is based on the following three typical situations in a site. The situations are chose based on the potential conditions in a brick façade site survey.
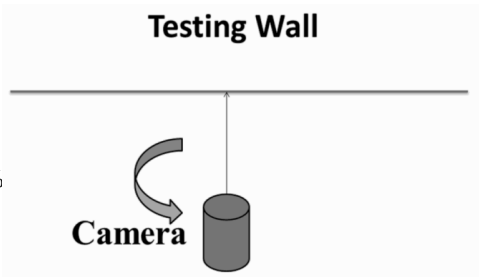

Figure 4 Camera rotation

We tested the algorithm on 12 images or video frames. In the site, workers may take the video for the whole façade in various locations and tilted angles because of obstacles or difficult site conditions. This case is performed by rotating the camera while videotaping.

Different rotation angles were tested. The setup of the equipment used and brick façade is shown in figure 4.

The validation is done by comparing the results of the algorithm with manual counts.

Both images and videos were taken on the same day under sufficient light condition. Figure 5 depicts an example of brick detection taken in a sloped location. The precision is 97% and the recall is 88%. This case shows that, the brick counting method would be able to distinguish bricks from the background.

Figure 6 is an example of 80 degree anticlockwise tilting of the camera. The precision is 99% and the recall is 92%. The result shows that in this case the tilted angle didn't affect the results.
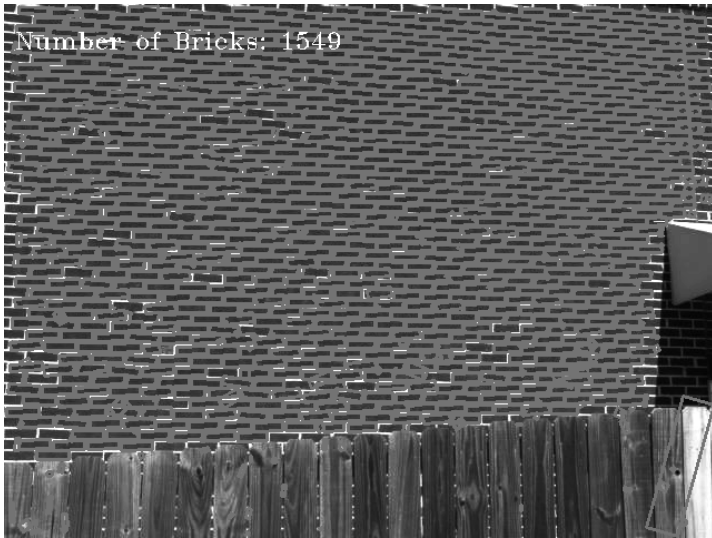
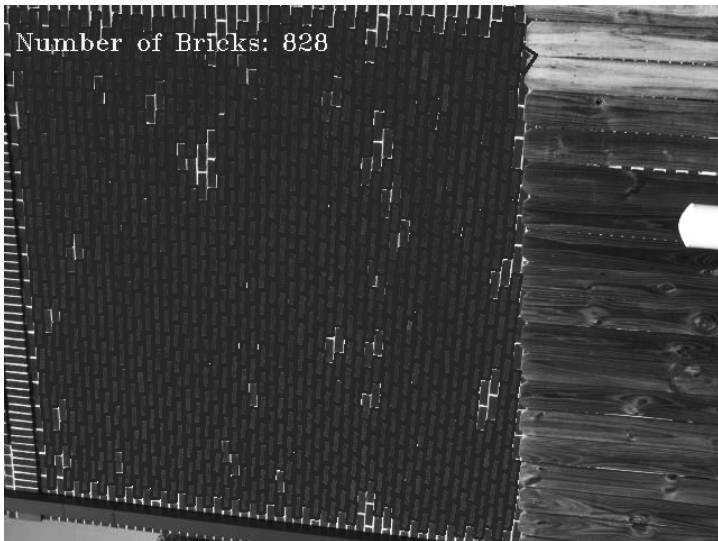**Figure 5   Brick detection tilted angle 45°**

Figure 6   Brick detection tilted angle 80°

The overall average of precision is 98%, and the overall recall is 90%. The time required for the detection process is less than one second.

From the above result, the confidence interval of precision is 96% to 99%, and the confidence interval of recall is 88% to 93%. The results show that the proposed method can perform robustly under different situations.

CONCLUSION

The proposed method uses the images obtained by a portable camera with relatively low cost to determine the number of bricks in a wall. The bricks are distinguished by their special properties including the following: rectangular, equal sized and red colored. The method has been tested on a variety of images taken from different angles in different directions and distances.

The results of the experiments show very high precision, with the overall precision being over 98%, indicating that the algorithm can detect the bricks accurately. The overall recall is 90%, showing that the algorithm is able to detect most of the bricks in the image. The algorithm is not affected by the rotation of the camera, and the percentage of the area of the bricks included in the image.

The results validated the ability of using visual data to automatically count the number of bricks on a brick façade. The brick count method suggested is able to accurately estimate the number of bricks on a wall. Future work will be focusing on generalizing to other brick structures, and to count more bricks on adjacent façade images.

ACKNOWLEDGEMENTS

REFERENCE

Assaf, S. A., and Al-Hejji, S. (2006). "Causes of delay in large construction projects." *International Journal of Project Management*, 24(4), 349-357.

Junji, H., Li, S., Jianli, X., Jun, C., and Ying, Z. "Size Detection of Firebricks Based on Machine Vision Technology." *Proc., Measuring Technology and Mechatronics Automation (ICMTMA), 2010 International Conference on*, IEEE, 394-397.

Schmitt, K., Young, R., Riddington, J., Budgett, D., and Chatwin, C. (2000). "Image processing applied to brick quality control." *The International Journal of Advanced Manufacturing Technology*, 16(6), 434-440.

United-States-Census-Bureau (2012). "Construction Progress Reporting Survey (CPRS) ", <http://bhs.econ.census.gov/bhs/cprs/about.html>.

Vermont-Government-Website (2003). "Reducing Vermont's Construction Waste ", <http://www.anr.state.vt.us/dec/wastediv/recycling/CandD.htm>.

Representation of Damage Information for Post-Earthquake Damage Assessment of Reinforced Concrete Frames

E. B. Anil[1], B. Akinci[2], J. H. Garrett[3], O. Kurc[4]

[1]Carnegie Mellon University, Department of Civil and Environmental Engineering, 5000 Forbes Ave., Pittsburgh, PA, 15213, eanil@andrew.cmu.edu
[2]Carnegie Mellon University, Department of Civil and Environmental Engineering, 5000 Forbes Ave., Pittsburgh, PA, 15213, bakinci@cmu.edu
[3]Carnegie Mellon University, Department of Civil and Environmental Engineering, 5000 Forbes Ave., Pittsburgh, PA, 15213, garrett@cmu.edu
[4]Middle East Technical University, Department of Civil Engineering, Ankara, Turkey, 06800, kurc@metu.edu.tr

ABSTRACT

Current practice of post-earthquake damage assessment is based on visual inspections by experts. Research has shown that visual inspection and reports produced as a result of inspections can be subjective, incomplete, and vary in terms of the level of detail and thoroughness. Critical data that provides information about damage can be captured using laser scanners and processed for a more objective and information driven damage assessment practice. In addition to damage information, the damage severity identification process requires reasoning about the structural behavior and topology. Building Information Modeling can be used as the underlying information source to streamline the damage assessment process. This research aims at representing crack information using a Building Information Modeling approach. In order to support reasoning for damage assessment, three activities were performed: 1) requirements were identified based on accepted damage assessment codes and standards; 2) capabilities of Industry Foundation Classes were explored for representing crack information; and 3) suggestions were provided for using existing concepts in IFCs for representing cracks.

Keywords: Building Information Modeling, Industry Foundation Classes, Crack, Earthquake Damage Assessment

INTRODUCTION

Current practice of detailed assessment of damage to buildings after earthquakes requires manual inspection of buildings and interpretation of inspection results by experts (FEMA, 1999; Menches et al., 2008). The results of inspections are often in the form of hand sketches showing building layouts, locations of inspected components in a building, and observed damages on the components. Experts manually interpret these sketches to identify the damage modes and damage severities of components (FEMA, 1999).

Interpretation of inspection reports and sketches and identifying damage severity requires expertise and understanding of the structural interaction between components (e.g., beams, columns, and walls making up an assembly). However, several challenges exist in the process of identifying damage severity. First, damage severity is not only a function of the observed damage (e.g., crack widths). Observed damage is useful only if the damage mode (e.g., shear, flexure, etc.) is known. More precisely, the same crack width and orientation can correspond to different damage severities for two different behavior modes. Second, identifying damage modes of components requires understanding how different components composing an assembly interact to withstand to earthquake forces. In the current practice, this is an informal thought process based on expert knowledge (FEMA, 1999).

Visual observation, manual reporting and interpretation have been shown to have drawbacks (Kempton et al., 2001; Menches et al., 2008; Phares et al., 2001). First, important details can be missed, misinterpreted, or field measurements and calculations can be incorrect during the inspection. Second, visual inspections and reports may not be as thorough as required. Some factors affecting the detailing and thoroughness of the inspection and the reports are safety concerns of the inspectors, experience and training, and time constraints. Finally, interpretation of the reports and overall results of the inspection can vary (Kempton et al., 2001; Menches et al., 2008; Phares et al., 2001).

In order to address these challenges, we need objective and accurate methods of collecting, storing, and processing damage information. Laser scanners have been shown to be efficient for detecting cracks and other damage indicators due to their long range, light independence, and accuracy (Figure 1) (Anil et al., 2013). Additionally, for a detailed assessment, the quantities and details of reinforcement bars, material properties, connectivity of components are needed (FEMA, 1999). Therefore, we need the semantics of the building components, as well as mechanisms for processing the damage information and structural properties of the building to

identify the damage severity of the buildings. Building information models can support representation and querying of the required properties of building components as well as topological relationships between components (Paul and Borrmann, 2009). However, current BIM approaches do not support representation of damage information.

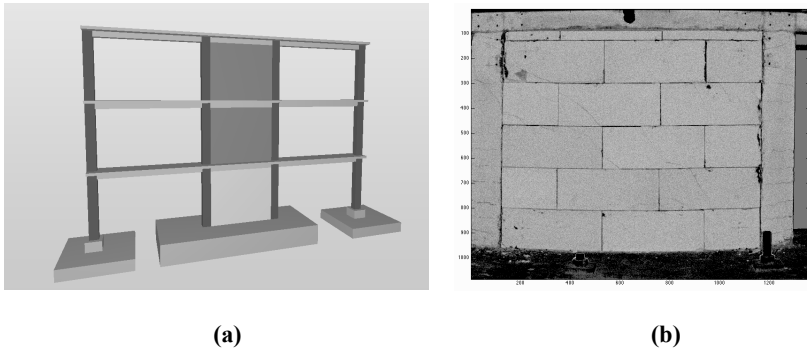(a)   (b)

Figure 1 IFC of a three-story three-bay test structure. The frame was subjected to lateral earthquake forces. The infill wall in the first story was laser scanned in the damaged state (highlighted wall in (a)). The intensity image shows several cracks (b).

This paper focuses on representing crack information captured using laser scanners to support identifying damage severities of components. We explored several methods of representing important crack information using existing capabilities of Industry Foundation Classes (IFCs). Only basic properties of cracks, such as width, location, orientation, and path were considered. These properties can be detected and measured by automated algorithms (Tsai et al., 2012). Utilization of texture maps and association of raw crack information to building components through the usage of external documents were discussed.

As a part of a larger project, we are working on: 1) representing all of the damage indicators required for damage assessment (i.e., cracks, spalling, permanent deformations, crushing, buckling and fracture of reinforcement); and 2) developing reasoning mechanisms, which can operate on these models to identify damage severity. Design and implementation of these reasoning mechanisms, and representation of other damage indicators are ongoing research tasks. This paper presents the preliminary results on the representation of raw crack information within existing BIMs.

REPRESENTATION REQUIREMENTS AND RESEARCH APPROACH

In the current practice, cracking constitutes the primary visual damage cue for assessment. Generally, damage initiates as cracking. Especially at insignificant to moderate damage severities, cracking is often the most important damage cue. As the damage advances, other damage cues add onto cracking, such as spalling of cover, which exposes reinforcement bars, buckling of reinforcement, and crushing of core concrete and infills, splitting of the cover concrete due to lap splice slips (FEMA, 1999).

An engineer uses relative location (relative to the component geometry and to each other), orientation, extent, and width of cracks on a component to classify damage (FEMA, 1999). For example, on a vertical component (e.g., column), horizontal cracks indicate a flexural type of damage, whereas diagonal cracks indicate a shear type of damage. Vertical cracks along the sides of the components may indicate splitting of the cover concrete due to slipping of lap splice. Horizontal cracking along the bottom of a component may indicate a sliding plane. Concentration of cracks at the corners of an infill wall may indicate crushing. Widths of cracks indicate the severity of the damage for different damage modes.

Several representation schemas can be possible depending on the usage and the reasoning mechanism, which operates on the model. For example, it is possible to derive some fundamental features using crack patterns, orientation, and width to be used with classification algorithms (Tsai et al., 2012). Regardless of the end features derived from the detected cracks, detection of the crack location, path, and width is often the first step of crack classification algorithms (e.g., Tsai et al., 2012).

The purpose of this study was to explore methods of representing the basic crack information, which can be detected by crack detection algorithms from laser scan data. Specifically, we are interested in representing the crack path, width, and relative location of cracks on a given component surface (Figure 2). Such, a representation is similar to the current practice of preparing sketches of building components and observed cracks.

Industry Foundation Classes currently do not support representing damaged conditions of buildings. However, several other mechanisms, which exist within IFCs, could be used for representing some of the required crack information. Minimal modification is desired to extend the usage of IFC (Akinci and Boukamp, 2002). Therefore, an appropriate strategy is extending IFC only when existing concepts are not adequate, and to introduce new concepts only when current representations or extensions cannot satisfy the requirements (Akinci and Boukamp, 2002).
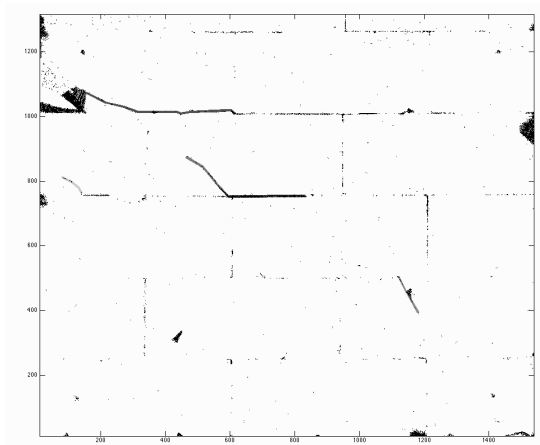
Figure 2 Results of the point classification and crack polylines (colored lines). Points are classified as being a surface point or a crack point using a voting method. In the figure, points classified as belonging to the surface are hidden. Cracks are traced manually.

We considered both the latest IFC release in development (2x4 RC3) and the latest released version (2x3) in this study. The reason is that several objects of 2x3, which are related to the concepts discussed in this paper, have been deleted in 2x4. However, 2x4 is not released as of the writing of this paper. These changes do not affect the conceptual discussions in this paper. Differences between the two versions are highlighted.

For detecting the cracks, we used a 3D implementation of Anil et al. (2013) to classify laser scan points as cracks or surface points (Figure 2). In this method, a plane is fitted to a region of fixed size around every point using a robust method, such as RANSAC, and a voting method is used to classify points. In every local region, if the number of points that are within 2σ distance to the local surface are greater than the number of points that are farther than 2σ distance, then the point is classified as being a crack point. Here, σ indicates the standard deviation of range measurements. We used a neighborhood size of 10 cm in this example. This method is sensitive to other off-surface points, such as mortar between bricks (Figure 2). The actual cracks in Figure 2 are traced manually.

REPRESENTATION METHODS

Following the minimal modification strategy to IFCs, we identified two applicable concepts for representing basic crack information. These are:

1- Texture maps to link crack information to component surfaces using presentation methods in IFC.
2- Storing crack information external to the model and linking the external file to components using *IfcRelAssociatesDocument* object.

Texture Maps

IFC supports annotations, such as text, hatching, tiling, coloring, and texture mapping. These annotations can be attached to semantic objects or can be free, such as grid lines. IFC allows representing geometric representations and styling information

Texture maps, in particular, can be in the form of 2D images attached to surfaces. A coordinate mapping exists between the 2D image space and 3D space. Styling information is attached to surfaces using the *IfcStyledItem* object (Figure 3). Using the texture mapping mechanism, images containing the crack information can be mapped to component surfaces.
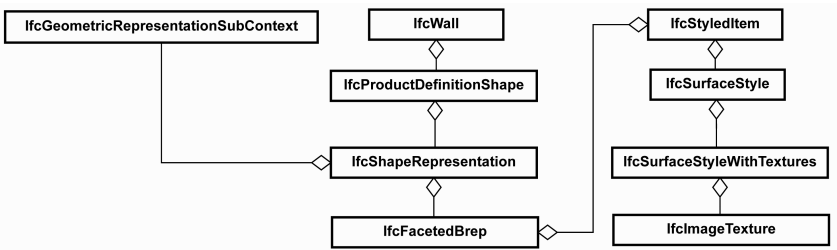

Figure 3 UML class diagram of objects related to texture mapping.

IFC allows representing and controlling the appearance of components under different contexts and views using the *IfcGeometricRepresentationContext* and its sub-class *IfcGeometricRepresentationSubContext* (Figure 3). The TargetView attribute of the sub-context object allows user-defined views. The UserDefinedViewTargetView attribute can be set to "Crack_Map" or any other unique identifier to distinguish the crack information from other styling. However, developers and implementers should agree on the usage and possible values for such an implementation to be useful in practice.

Some of the entities used for texture mapping in IFC 2x3 are deleted in IFC 2x4. Among these are *IfcStyledPresentation* object, which is used for linking the *IfcStyledItem* to *IfcSurfaceStyle,* and *IfcAnnotationSurface* object, which is used to represent surfaces with texture coordinates assigned.

External Documents

Crack information can be stored in external documents and can be linked to building components using *IfcRelAssociatedDocument* relationship. A single document can be linked to multiple components. One potential benefit of linking external documents is that the raw point cloud can be linked to components in addition to processed information. Raw point cloud can be useful for later processing, if necessary.

In this approach, the crack information can be stored in external text files, where a crack is represented with a polygon approximating the crack path and a vector of crack width measurements along the crack path (Figure 4). Similarly, the raw point cloud can be linked to the components.

CrackFile.txt

#Crack_1			
#X	Y	Z	Crack_Width
-2.4504759	-2.9318161	-0.94557101	1.22
-2.4570391	-2.9396689	-0.94374001	1.22
-2.4610839	-2.94451	-0.94102699	1.00
-2.463773	-2.9477291	-0.93771899	1.00
#Crack_2			
…			

Figure 4 Detected crack paths and widths can be stored in external files as polylines. Each line is the x, y, z coordinates of a polyline in the project coordinates and the measured crack width at that vertex of the polyline.

CONCLUSIONS AND FUTURE WORK

This paper presents a study that was conducted to explore how to represent basic crack information detected by laser scanners. The crack properties represented are crack path and width. By minimally changing the IFC specification we found two potential methods of representation. One method is representing crack information texture maps. The second method is externally linking crack information to building

components. These two methods require minimal change in the IFC schema to support representation of crack information linked to building components.

Future research is required to validate whether the two representation methods suggested in this paper can support identification of damage severities of components or whether more formal representation of cracks is needed within IFCs to streamline and possibly automate the reasoning about cracks. Other future research directions are developing reasoning mechanisms to support damage assessment, and representation of other damage indicators, such as spalling, fracture of reinforcement and crushing.

REFERENCES

Akinci, B., & Boukamp, F. (2002). Representation and integration of as-built information to IFC based product and process models. *ISARC*.

Anil, E. B., Akinci, B., Garrett, J., & Kurc, O. (2013). Characterization of laser scanners for detecting cracks for post-earthquake damage assessment. *ISARC 2013*.

FEMA. (1999). *FEMA 306 - Evaluation of Earthquake Damaged Concrete and Masonry Wall Buildings: Basic Procedures Manual*. Washington, D.C.: FEMA.

Kempton, J., Nichol, S., Anumba, C., & Dickens, J. (2001). Surveyor variability in large-scale house condition surveys. *Structural Survey*, *19*(4), 156–162. doi:10.1108/02630800110406658

Menches, C. L., Markman, A. B., & Jones, R. J. (2008). Innovative Method for Investigating the Facility Damage Assessment Process. *Architectural Engineering Conference (AEI) 2008*. Denver, Colorado.

Paul, N., & Borrmann, A. (2009). Geometrical and topological approaches in building information modeling. *ITCON*, *14*, 705.

Phares, B. M., Rolander, D. D., Graybeal, B. A., & Washer, G. (2001). Reliability of visual bridge inspections. *Public Roads*, *64*(5).

Tang, P., Huber, D., Akinci, B., Lipman, R., & Lytle, A. (2010). Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques. *Automation in Construction*, *19*(7), 829–843. doi:10.1016/j.autcon.2010.06.007

Tsai, Y.-C., Jiang, C., & Huang, Y. (2012). A multi-scale crack fundamental element model for real-worl pavement crack classification. *Journal of computing in civil engineering*, *Accepted m*. doi:10.1061/(ASCE)CP.1943-5487.0000271

SketchIT: A Sketch-Based Modeling Environment for Structural Analysis

Li Ge[1] and Falko Kuester[1,2]

[1]Department of Structural Engineering, University of California,
San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0085; email: lge@ucsd.edu
[2] California Institute for Telecommunications and Information, University of
California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0085; email:
fkuester@ucsd.edu

ABSTRACT

Computational techniques for structural analysis have been systematically studied and developed over the years, while modeling interfaces for these environments have remained largely unchanged. Typically, a so-called WIMP (Windows, Icon, Menu, Pointer) or a script-based interface is used. Even though easy to implement, these interfaces are frequently counterintuitive for the task at hand, abstracting natural hands-on modeling tasks into point, click and drag events. As a result, even experienced engineers have to spend large amounts of time to translate their models into the form that computer can recognize. With the emergence of new input devices such as digital pens and multi-touch enabled displays, a new generation of sketch-based input technique has become possible. This paper presents a modeling and simulation environment for structural analysis using sketch-based input. Two case studies including the static analysis of a 2D elastic frame and transient analysis of a 2D nonlinear frame are presented, showing that model creation, simulation and analysis can seamlessly co-exist, supporting the creation of an integrative analysis methodology.

INTRODUCTION AND RELEATED WORK

Sketches are a common way for engineers to share conceptual ideas preceding the use of WIMP (Windows, Icon, Menu, Pointer) based engineering software for the creation of a formalized design. With only a few strokes, complex shapes can be evoked, creating a natural and intuitive setting to communicate and explore concepts.

Sketch recognition systems are being developed for many hand-drawn diagrammatic domains such as mechanical engineering (Alvarado, 2000; Masry & Lipson, 2007), Unified Modeling Language (UML) class diagrams (T Hammond & Davis, 2006), web-page design (Lin, Newman, Hong, & Landay, 2000), GUI design (Caetano, Goulart, Fonseca, & Jorge, 2002), virtual reality (Do, 2001), and many others.

In structural engineering, the use of sketches during the initial design phase is even more likely, followed by computer-aided modeling and simulation-based

analysis. As such, combing sketching, modeling, simulation, and analysis in a single, integrative structural engineering environment, holds great promise. Hutchinson et al. (Hutchinson, Kuester, Phair, & others, 2007) introduced a sketch-centric 2D FEA approach, focusing on the static analysis of 2D elastic structural models. SketchIT takes this idea further by integrating sketching, sketch recognition and model construction with support for static, nonlinear and dynamic finite-element analysis, and visualization. SketchIT supports the intuitive and interactive creation of non-structural models coupled with immediate-mode simulation, allowing simulated and experimental data to be co-located and analyzed in unison.

TECHNICAL APPROACH

SketchIT was designed using a cloud-centric approach, decoupling its graphical user interface from the sketch, gesture and multi-touch detector, finite element model builder, solver and results viewer (Figure 1). Specifically, the underlying client-server architecture (Berson, 1996) distributes the graphical user interface and computationally intensive components across different devices. The sketch recognizer, the model builder and the graphics viewer are deployed on the client side, while the FEA solver, the Tcl interpreter and a Websocket interface reside on the server side. Its web browser enabled graphical user interface, allows SketchIT to run on any device supporting a standards compliant browser.
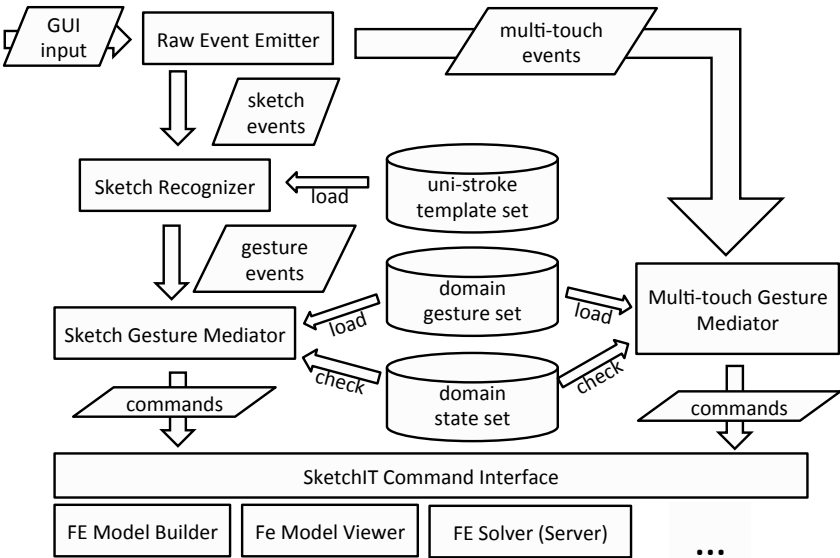

Figure 1: SketchIT modeling, simulation and visualization pipeline

SKETCH-BASED MODEL BUILDING

The SketchIT approach is particularly suitable for touch sensitive input devices, such as tablets and smartphones, supporting the use of natural sketch gestures. On traditional desktop systems, sketch gestures may be obtained through external digitizer tables or simulated using keyboard and mouse events. Standard multi-touch gestures such as two-finger pinching, two-finger swiping and two-finger rotating can be used to zoom in, translate and rotate the workspace and the processing pipeline is conceptually summarized in Figure 1.

Gesture recognition. The sketch gestures are defined by the combination of primitive shape types, domain related constraints and actions to be triggered (Tracy Hammond & Davis, 2005). SketchIT's visual language is summarized in Figure 2 and a collection of basic sketch gesture examples is shown in Figure 3, illustrating that a compact and intuitive description can be created. The separation of shape primitives and domain specific gestures makes this approach flexible and extensible.

Primitive		State	Action
Line			Add a 1D element block(truss/beam)
Triangle		First point near a node	Add a single point constraint
Circle		Center near a single point constrained node	Add a pin
Circle		Center near triangle edge	Add a roller
Arrow		Head or tail near a node	Add a nodal load
Arrow		Head or tail far away from any nodes	Add uniform loads
Delete			Remove selected components
Cancel			Cancel selection
Check			Run analysis
Pigtail_1			Undo
Pigtail_2			Redo
Letter_m			Mesh selected element blocks
Open curve		Select mode is activated	Select components that intersect with the curve
Close curve		Select mode is activated	Select components that are inside the curve
Path		Move mode is activated	Move selected object by an arbitrary path

Figure 2: Visual language consisting of sketch primitives, states and actions

Sketch recognition is decomposed into three stages, including:

Shape Recognition. With the focus on 2D frame structures, the primitive shape set can be compact and the computational cost of matching shapes in the templates library moderate. For improved interactivity, the shape recognizer resides on the client side and utilizes the $1 Recognizer algorithm (Wobbrock, Wilson, & Li, 2007), a light-weight sketch recognition algorithm with high accuracy. Users are asked to draw symbols, defined by the vocabulary, in one stroke. Once a stroke is finished, the system notifies the recognizer, which in turn takes the strokes sequence as input, and generates the matched template name. The template is predefined by its geometry properties and loaded by the recognizer.
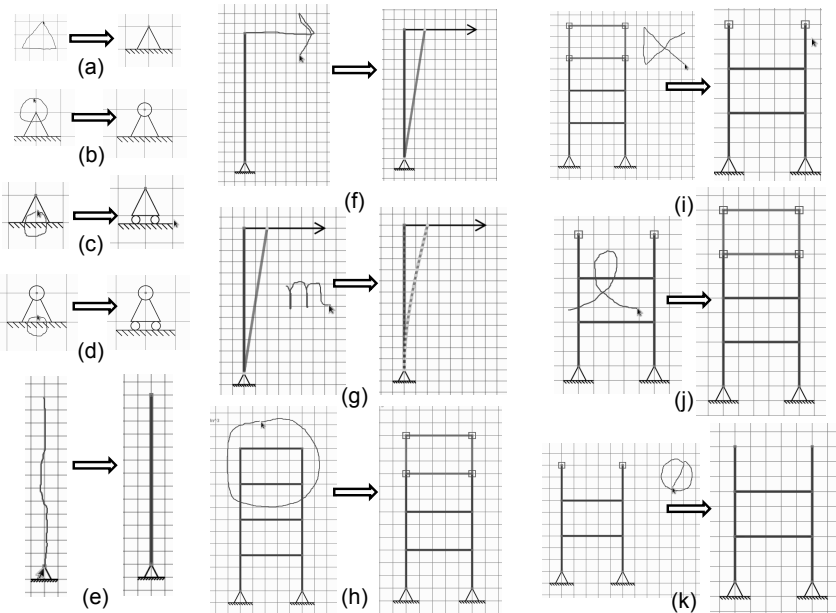
Figure 3: Examples of basic sketch gestures to construct constraints (a)-(d), construct a beam (e), apply a load (f), refine meshing (g), select part of the model (h), delete a member (i), undo a previous operation (j) and deselect nodes (k).

Post feature extraction. Once the shape recognizer has obtained the primitive shape type, it is possible to extract the feature parameters based on the knowledge of recognized shape types for higher-level gesture recognition. For example, if the recognized shape is of type "circle", then feature properties radius and center location will be add to the recognized object.

Gesture-to-command mapping. A gesture template is defined by its primitive shape and domain specific constraints. After the first two steps described above, a recognized shape object with feature parameters is produced. It provides a clear interface for the domain modules to perform a domain command. The mappings between these gesture objects and domain commands are encapsulated and maintained in a mediator object.

Geometry Description. The system keeps a collection of geometric cells to store the geometric information of the model. A geometric cell is defined its topological and metric information. Topological information indicates the connectivity of the geometric cell, i.e., how the geometric cells connect to the points in the space, and is managed by a snap procedure. The feature so-called 'object snap' is now a standard functionality of most CAD software, including AutoCAD (Onstott, 2013), SolidWorks (Onwubolu, 2012), etc. The metric information includes the coordinates of the control points and the shape interpolation function of the geometric cell. The metric information is obtained by applying a linear transformation to the screen coordinates. Users can customize the origin of the visualized coordinate

system and set the unit length of the grid, i.e. the mapping between displayed and real-world distances.

Element Description. SketchIT supports four types of line elements, which are commonly used in frame structures, including (1) elastic beam-column elements, constructed with two nodes and have three degree-of-freedom (DOF) per node; (2) truss elements, constructed with two nodes and one degree of freedom (axial) per node; (3) Nonlinear beam-column elements, constructed with two nodes, with three DOF per node; (4) Beam with plastic hinges, constructed with two nodes, with three DOF per node. Element may be selected through using one of two selection gestures implemented under the 'select' context, 1) any elements completely within the region defined by a sketched polygon will be selected; 2) any elements that are intersected by a stoke will be selected. Once elements are selected, a calculator-like menu panel will automatically open, allowing properties of selected elements to be edited.

Model Discretization. Beam and truss elements can be automatically discretized to improve the accuracy of results. After setting the maximum length of a line element (default is the grid space) in the mesh setting menu, SketchIT will perform the discretization automatically. This process is shown in Figure 3 (g).

Boundary Conditions and Loads. In a 2D frame system, a node has three degree of freedom., translation along horizontal direction, translation along vertical direction and rotation in the plane. There are four types of single point constraint as shown in Figure 3 (a) – (d). Two types of loads, i.e., nodal loads and uniform loads, are supported. Both of them are applied by sketching an arrow gesture as shown in Figure 3 (f). If the arrow is snapped to an existing node, it will be recognized as a nodal load. Otherwise it will be recognized as a uniform load and applied to all existing nodes. For the nodal loads, the default direction and magnitude are determined by a user drawn arrow symbol.

Loading Time Histories. In dynamic analysis, the loading time history needs to be specified. SketchIT considers four common time history types: linear, constant, sine and triangle. SketchIT also allows time history data to be imported via a user defined text file.

SIMULATION & VISUALIZATION

Once the model has been created and boundary conditions sufficiently defined, it is automatically forwarded to the FEA solver, running on a remote server. The FEA solver performs the needed computation and streams the field data (nodal displacement, element forces, etc.) back to the requestor for result visualization. In small deflection theory, and in general for most FEA problems, the actual displacements calculated are very small compared to the size of the model. So for the purpose of visualization, the displacements need to be exaggerated. Results are automatically scaled, such that the maximum displacement or translation (either vertical or horizontal) is clearly visible. By default this is configured to reach 40 on-screen pixels to be clearly discernable. Users can override the scale factor and change it according to their preferences. For dynamic analysis, the FEA server will continue to deliver field data such as nodal displacements to the client. The client subsequently creates the corresponding animation, caching just enough data points to allow for

smooth simulation. Resulting in near real-time response. An animation widget (Figure 5 b) can be used to control the animation.

IMPLEMENTATION

The client side was implemented in JavaScript, a highly flexible scripting language for web-browsers, with GUI (Graphical User Interface) drawing from Sencha Touch (Sencha Inc, 2012) for multi-touch interaction support. The server side FEA solver is based on the Open System for Earthquake Engineering Simulation, OpenSees (Mazzoni, McKenna, Scott, Fenves, & others, 2005), known for its openness and extendibility and the needed interface provided through the Tcl scripting environment (Ousterhout & Jones, 1994). Websocket (Fette & Melnikov, 2011) technology is used for client-server communication.

CASE STUDY

Two case studies, the static analysis of a 2D elastic frame and transient analysis of a 2D nonlinear frame are presented next.

Static analysis of a 2D elastic frame is illustrated for a 3-story frame structure created from a sketch. The workflow is shown in Figure 4. Notice that the sketches are automatically beautified after being recognized. All structural components of the frame are modeled by 2D elastic beam elements. Element property configurations are done through a calculator-like interface as shown in Figure 4 (a). In this example, all beams are of same section property, i.e., cross-section area **A** is 44.8 in^2, moment of inertia **I** is 8160 in^2 and Young's modulus **E** is 29000 klbf/in^2. For static analysis, immediate mode analysis is performed, when the first load is applied, resulting in the model being discretized using the current grid size, the model being forwarded to and simulated on the server, before deformation results are visualized, side-by-side with the sketched model. Whenever a change happens in the system, such as loads, boundary conditions, element properties, being modified, the simulation and visualization updated. This feature provides a very intuitive way to acquire conceptual knowledge of the system, allowing users to receive feedback immediately after a model change event occurs (Figure 4 b to i).

Transient analysis of a 2D nonlinear frame is illustrated for a 3-story frame structure created from a sketch. The structural beams and columns are modeled using 2D force-based non-linear beam elements. For the beams, five sections are used for each element and a wide flange steel fiber section used for both beams and columns. The boundary condition settings are the same as in the static cases. The uniform excitation load is added by using an arrow gesture. The default time history data for uniform excitation loading is from El Centro, Imperial Valley, California earthquake on May 18, 1940. The dynamic analysis is trigger by drawing a check gesture (Figure 5 a). Once the analysis starts, the FEA solver on the server side will continuously update and stream result from each analysis step to the client, where it is visualized as a continuous animation.
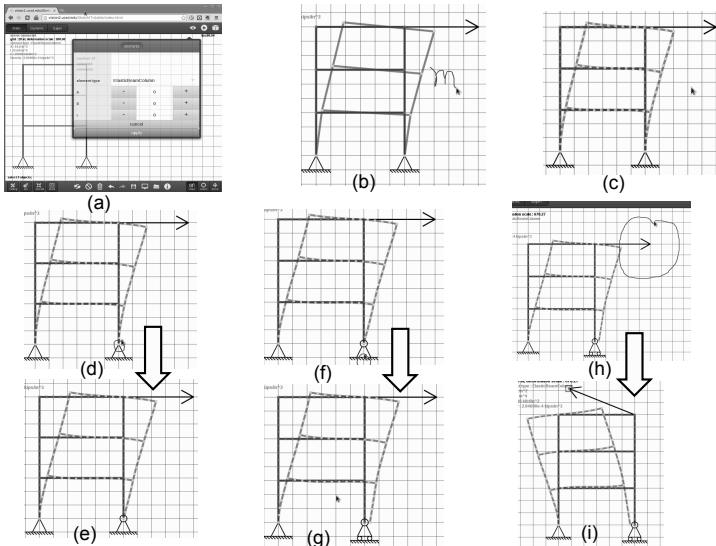
Figure 4: static analysis of 2D frame, (a) creation of model and definition of boundary conditions, (b) gesture initiated remeshing and (c) immediate results update, (d) – (g) modification of constraints, (h) – (i) change of load direction and magnitude.
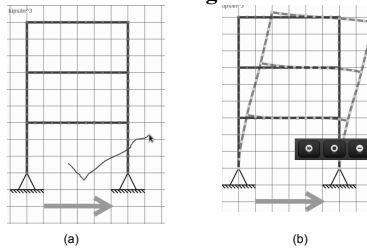
Figure 5: Dynamic analysis of 2D nonlinear frame, (a) triggered by a checkmark gesture and (b) interactive visualization of simulation results.

CONCLUSION

A 2D sketch-based, integrative structural analysis environment, termed SketchIT, is presented. SketchIT integrates a sketch-based interface, finite element analysis, visualization and analysis, into a web-based workspace. SketchIT supports static, linear, as well as dynamic and nonlinear structural analysis. Two case studies are presented, demonstrating that the sketch-based modeling and analysis approach is intuitive and efficient. A limitation of the system is it currently only support 2D modeling. The future improvement includes the extension of this system to three dimensions.

ACKNOWLEDGMENT

REFERENCES

Alvarado, C. (2000). *A natural sketching environment: Bringing the computer into early stages of mechanical design*. Massachusetts Institute of Technology.

Berson, A. (1996). *Client/server architecture*. McGraw-Hill, Inc.

Caetano, A., Goulart, N., Fonseca, M., & Jorge, J. (2002). Javasketchit: Issues in sketching the look of user interfaces. *AAAI Spring Symposium on Sketch Understanding* (pp. 9–14).

Do, E. Y. L. (2001). VR sketchpad. *Proceedings of the CAAD Futures Conference* (pp. 161–172).

Fette, I., & Melnikov, A. (2011). The WebSocket protocol.

Hammond, T, & Davis, R. (2006). Tahuti: A geometrical sketch recognition system for uml class diagrams. *ACM SIGGRAPH 2006 Courses*, 25.

Hammond, Tracy, & Davis, R. (2005). LADDER, a sketching language for user interface developers. *Computers & Graphics*, *29*(4), 518–532. doi:10.1016/j.cag.2005.05.005

Hutchinson, T. C. C., Kuester, F., Phair, M. E. E., & others. (2007). Sketching finite-element models within a unified two-dimensional framework. *Journal of computing in civil engineering*, *21*, 175.

Lin, J., Newman, M. W., Hong, J. I., & Landay, J. A. (2000). DENIM: finding a tighter fit between tools and practice for Web site design. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 510–517).

Masry, M., & Lipson, H. (2007). A sketch-based interface for iterative design and analysis of 3d objects. *ACM SIGGRAPH 2007 courses* (p. 31).

Mazzoni, S., McKenna, F., Scott, M. H. H., Fenves, G. L. L., & others. (2005). OpenSees command language manual. *Pacific Earthquake Engineering Research (PEER) Center*.

Onstott, S. (2013). *AutoCAD 2013 and AutoCAD LT 2013 Essentials* (p. 400). John Wiley & Sons.

Onwubolu, G. (2012). *A Comprehensive Introduction to SolidWorks 2012* (p. 912). SDC Publications.

Ousterhout, J. K., & Jones, K. (1994). *Tcl and the Tk toolkit* (Vol. 221). Addison-Wesley Reading, MA.

Sencha Inc. (2012). Sencha touch. Retrieved January 1, 2013, from http://www.sencha.com/products/touch

Wobbrock, J. O. O., Wilson, A. D. D., & Li, Y. (2007). Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. *Proceedings of the 20th annual ACM symposium on User interface software and technology* (pp. 159–168).

Design and Development of SAVES: A Construction Safety Training Augmented Virtuality Environment for Hazard Recognition and Severity Identification

Ao Chen[1], Mani Golparvar-Fard[2], and Brian Kleiner[3]

[1] Ph.D. Student, Vecellio Construction Engineering and Management Group, Via Department of Civil and Environmental Engineering and Myers-Lawson School of Construction, Virginia Tech, Blacksburg, VA; PH (540) 494-0202; FAX (540) 231-7532; email: aochen@vt.edu

[2] Assistant Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL; PH (217) 300-5226; FAX (217)333-9464;email: mgolpar@illinois.edu

[3] Director, the Myers-Lawson School of construction and Professor of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA; PH (540) 231-6777; FAX (540) 231-7532; email: bkleiner@vt.edu

ABSTRACT

One of the most challenging aspects of safety for construction sites is ensuring that workers can predict, identify, and respond to potential hazardous conditions before they are exposed. From a scientific standpoint, we currently lack the knowledge of discovering the most efficient training styles for safety and also understanding why and how these styles of training can influence the post-training activities. To address these needs, a System for Augmented Virtuality Environment Safety (SAVES) is designed and presented in this paper. SAVES which integrates a BIM with 2D images on a jobsite, allows trainees to conduct a set of interactions with SAVES and accomplish multiple instruction-based and task-based training modules. These modules include detection of ten types of hazard and/or energy sources at three levels of severity. The complete process of design, development and implementation of SAVES is demonstrated and the lessons learned are discussed in detail.

INTRODUCTION

Every year in the United States, numerous work-related fatalities and non-fatal injuries have happened in construction sites. 721 fatalities and 71,600 nonfatal accidents in construction were reported in 2011 (CDC, 2012). Most of them were caused by the unsafe actions and operations due to failure of following safety rules and guidelines. Many accidents could be prevented and minimized if

workers were able to take proper actions by rapidly identifying hazard types and their severity levels. Federal agencies have emphasized the importance of training program in reducing the injuries number in construction industry: "*develop*, *implement* and *enforce* a comprehensive safety and health training program in language and literacy level for workers, which include training in hazard recognition and the avoidance of unsafe conditions" (OSHA, 2009).

Safety awareness and hazard recognition in training are the keys to approach the goal of reducing unsafe practices. To that end, this paper proposes an application of game engine-based Augmented Virtuality (AV) which integrates Building Information Model (BIM) with 2D images as a viable form of safety training. This application can offer an interactive tool for training all construction workers on hazards and safe working procedures. The process involves developing a series of instructional-based and task-based AV modules that are intended to elevate safety awareness by explaining the different types of hazards and safety issues associated with their performance, and test the user's ability on re-solving problems while performing a required set of pre-defined tasks in the AV environment. Such a "learning by doing" opportunity creates a clear mapping between site hazards and their recognition and as a result can further enhance the user's expertise in recognizing the severity level of hazards.

BACKGROUND

To date, government and industry cooperate in developing diverse training programs in order to provide better safety culture. Mine Safety and Health Administration (MSHA) clearly indicated that training plays a critical role in preventing deaths, injuries, and illness on the job (MSHA,1998). Only with effective training can miners recognize possible hazards and know the safe procedures to follow. Unfortunately, among such training programs, many of them were not as effective as they could be in providing effective training. The common training methods such as lecture and safety videos only made 5% to 20% retention rate due to the limited two-dimension cognition design. Whereas, the three-dimension style, which was also known as "practice by doing" could bring at least 75% retention rate but rarely been used in safety hazard recognition. NIOSH had mentioned this necessity of improving training program that construction workers needed new training materials and videos or lectures could not be the only resources (NIOSH, 2002). Other industries like MSHA, already adapted VR as training tool for safety. The information and skills obtained from VR training could be transferred to the real world in a more expressive and realistic way than acquired by applying more conservative, didactic training

methods. Moreover, the biggest convenience was that VR permitted the trainees to experience situations that were hard or impossible to be shown in the real world. Another advantage of using VR was that it could systematically offer a wide range of possible training scenarios without suffering the high cost and risk of fielding personnel, equipment, and vehicles (Zeltzer et al., 1996).

Grant and Daigle (1995) had developed a VR environment to train the workers of power utility facility correctly operating the distribution device and switch controller. Lucas et al. (2008) had studied an applicable VR scenario on conveyor belt safety to help mine industry workers quickly and accurately operated buttons and switched without bringing any potential hazards. Despite the advantage of using VR in training, such applications didn't provide adequate freedom to users to experience different situations and they were lack of sufficient validations. Ku and Mahableshwarjer (2011) proposed a framework of BIM engaging with Second-Life to teach students integrated construction process. Such application lost realism in display and simply-inserted VR was away from real-world practice. Zhao et al. (2012) developed one safety scenario for electrician based on statistical results of accident reports. Such study presented a well-designed training stage with detailed conditions. But validation of correlation between designed training contents and raw data was not explained. Comparing with the natural drawbacks of Augmented Reality (AR) (e.g. impossible to put trainee to a real fully hazardous area, lack of interactive feedback) and VR (e.g. lose many realistic practice and display), Augmented Virtuality (AV) which have the advantages from both sides is proposed to answer the research questions. This research plans to examine the areas of safety in construction, hazard recognition, human centric AV development. The pedagogical goal of developing such AV environment focuses on providing some of the learning outcomes versus purely scenario driven learning outcomes such as lectures and safety videos.

DESIGN STRUCTURE OF SAVES

The purpose of developing SAVES is to improve skills of workers in identifying potential hazards and their severity in construction. As shown in Figure 1, SAVES is a highly renewable and expandable system. Clients provide specific training demands to SAVES. Research team designs proper BIM within SAVES and the initial training scenarios are set up for expert reviewing. The modified training points are employed by using SAVES database then finalized scenarios are wrapped up for official training. The final training package allows individual worker or groups experience SAVES either from on-site training room or their own laptops in anywhere. The training results and data are automatically

collected by SAVES and analysis will be sent back to SAVES research team and clients. Clients can easily check the training results and exam their employees' safety awareness and ability. Moreover, clients are able to see what safety knowledge and specific hazard awareness their employee are currently missing so that decision makers can change their initial training focus and let SAVES address these new focus in the next training cycle. Also, SAVES can extract different training scenarios from SAVES database and add more new contents to bridge different clients in order to build a comprehensive training program.
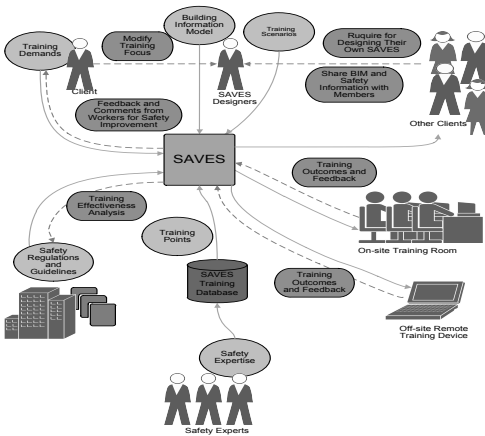
Figure 1. Framework of SAVES

DEVELOPMENT OF SAVES

The main research objective is to develop and implement an AV system to improve workers' awareness of hazard classification, severity identification and ability of taking proper reaction plan. To successfully complete SAVES, two key parts of training development are identified and explained.

Formalizing Hazard Inventory for Safety. Safety issues and potential hazards widely and randomly distribute around the whole construction site. It is hard to simply identify the causality from each injure or fatality since every case is unique and caused by the overlap of multi factors as well as other diverse variables under different conditions. In this research, hazard is defined as a condition or action that has the potential for an unplanned release of, or unwanted contact with, an energy source that may result in harm or injury to people, property, or the environment (Kleiner 2012, Chevron 2012).The research team goes through the documents, case studies, reports and has delivered top 10 energy sources types that most easily result in accidents and fatalities. The summary of identified 10 energy types are presented in table 1.

Developing Modules for Safety Training. One of the research objectives is helping trainee to rapidly recognize the severity level of hazard in order to take the proper action as well as reduce the loss of cost and damage. Inside SAVES database, nearly 600 hazard datum of major energy types are successfully decided and categorized. Around 300 hazard datum of sub-energy types have been recorded as well. Research team has developed a scale index system from 1 to 3

indicating different severity. The smaller the number the construction activity receives the better safety behavior it has. Level 1(green) presents no observable issues been presented in the training scenario. Level 2(yellow) shows that potential hazard or/and poor practice stated in the training scenario. The last level, level 3(red), requires an instant work stop to avoid foreseeable serious accidents. This color coded system can provide easy, directive signals in human information cognition process. All the information of such training elements are integrated together and those completed data are saved to our hazard information database for the use in final scenarios setup and other future research.

Table 1. Definitions of Energy Type for Safety Training

Energy Type	Definition	Energy Type	Definition
Biological	Living organisms that can present a hazard	Motion	The change in position of objects or substances
Chemical	The energy present in chemicals that inherently, or through reaction	Pressure	Energy applied by a compressed or vacuum liquid or gas
Electrical	The presence and flow of an electric charge	Radiation	The energy emitted from radioactive materials
Gravity	The force caused by the attraction of all other masses to the mass of the earth	Sound	A vibrating-cause force the energy is transferred through the substance in waves
Mechanical	The energy of the components of a mechanical system	Temperature	The measurement of differences in the thermal energy

IMPLEMENTING SAFETY TRAINING SCENARIOS IN SAVES

To emerge the designed safety training elements to the environment, a number of flash-based training scenarios have been developed and linked. These scenarios aid to demonstrate the usage of this research by letting trainees to interact in SAVES in a manner similar or even exact to conditions and situations as they would be on-site. Training scenarios are envisioned as small set exercises with an OSHA regulations and instructions inspired contents to evaluate trainee performance. Comparing with such instruction-based module, task-based training module requires trainees to use their safety knowledge to search for training

scenarios and elements then recognize the hazards types, severity levels and corresponding action standards. In addition, trainees will be tested to make a proper hierarchy control plan in each training scenario. SAVES is divided into 3 training stages with 68 hazard training scenarios (indoor-work environment, outdoor-work environment and construction equipment zone) by complexity and module types. This paper only illustrates developing the first training stage which focuses on training worker's general safety knowledge. All the training modules will be scored and shown to the trainee in screen in real time. Figure 2 shows the finalized design

Figure 2.Snapshot of SAVES and Training UI

of SAVES environment and UI of safety training scenarios inside SAVES. The weighting index and coefficient system, judge metrics, penalty system, validation methodology and detailed analysis of pilot research are not discussed and covered in this paper.

PILOT RESULTS

A qualitative and a quantitative analysis have been conducted in the pilot study. A survey is sent to the trainees who use SAVES in three different construction sites. All 36 participants answered survey and the statistic shows that 98% users have positive attitudes about SAVES and 100% users indicate that they are highly engaged with this AV. As for training effectiveness, 94% users agree the training contents that presented in SAVE and 97% users think they have more safety awareness and confidence after implementing SAVES. The effectiveness follow-up research will be studied and presented in the future.

The quantitative analysis is conducted based on the 36 sets of SAVES outcomes. The pilot results show the average accuracy of training is 84% and the average responding time is 1.5 minutes. However, as shown in table 2, among 36 users in total, they usually make incorrect discrimination with mechanical hazard (level2), chemical hazard (leve2), electrical energy (level1), sound energy (level1) and gravity energy (level1). This suggests the research participating companies should pay more particular attentions to train their employees in such weak points

in their next training cycle. The detailed research will be presented in the future with more available data.

Table 2. Incorrect Discrimination Probability in SAVES

	Bio	Che	Elec	Gra	Mech	Mot	Pres	Rad	Sod	Temp
Lv1	0.026	0.013	0.128	0.077	0.051	0.013	N/A	N/A	0.077	0.013
Lv2	0.013	0.064	0.013	0.026	0.115	0.051	0.013	N/A	0.051	0.026
Lv3	0.026	0.038	0.038	0.026	0.013	0.051	0.026	N/A	0.013	N/A

CONCLUSION

This research has designed and developed SAVES to study its potential as a training tool for the improvement of safety awareness and ability of hazard recognition in construction. Such practical exercises provide the most transferable knowledge and skills back to the trainee. This "learning by doing" style creates a clean mapping between site hazards and their recognition with short modeling time. The correlation allows practitioners to exercise recognition of site hazards more rapidly and in turn avoid fatalities and injuries through improvement of situational awareness.

The current pilot research shows that AV is able to provide efficient training effort since workers can quickly accept SAVES with little problem and their safety knowledge are enhanced by interacting with SAVES. Besides, the top 5 hazards with severity levels which having high ratio of wrong discrimination are identified. SAVES makes safety practical learning more active and engaging since it allows for safe simulation of real-life events in a digital environment that might otherwise be too dangerous or expensive. Construction workers, supervisors, owner representatives, contractors and society will benefit from such advantages.

The future work will focus on developing the weighting index and deciding the complexity coefficient for better training accuracy measurement. Also, the detailed analysis and discussion of research outcomes, studying the best way to maximize the benefits of SAVES and studying trainees' hazard tolerance will be the upcoming research focuses.

ACKNOWLEDGEMENT

REFERENCES

Bird, F.E. Jr., and Germain, G.R. (1992). *Practical Loss Control Leadership*. Institute Publishing, Loganville, Georgia.

Centers for Disease Control & Prevention (2012). *Construction.* [online] Available at <http://www.cdc.gov/niosh/construction/> [Accessed 15 May 2012]

Daniels, W. J., and Salisbury, S. A. (1999). IITRI's Simulator Prepares Train Operator. *R & D Magazine*, 41(99), pp.43-55

Garant, E., Daigle, A., Desbiens, P., et al. (1995). A Virtual Reality Training System for Power-Utility Personnel, *IEEE Pacific Rim Conference on Communication, Computers, and Signal Processing*, pp.296–299.

Ku, K and Mahabaleshwarkar,S, P. (2011). Building Interactive Modeling for Construction Education in Virtual Worlds, ITCon 2010, Vol.16, pp 189-208.

Lucas, J., McMahan, R., Engle, R., Bowman, D., Thabet, W., Scafrik, S, and Karmis, M. (2008). Improving Health and Safety through Conveyor System Training in a Virtual Environment, *International Future Mining Conference*, 9-11 June, 2008. Sydney, Australia.

Mine Safety and Health Administration (MSHA). (1998). *Educational Field Services, A New Focus in MSHA* [online] Available at:<http://www. msha.gov/TRAINING/EPDFS/EPDFS.pdf> [Accessed 23 March 2012].

National Institute for Occupational Safety and Health (NIOSH). (2002). *Recommendations to The U.S. Department of Labor for Changes to Hazardous Orders*, [online] Available at:<www.cdc.gov/niosh/docs/ nioshrecsdolhaz/pdfs/dol-recomm.pdf> [Accessed 18 March 2012].

Occupational Safety and Health Administration (OSHA). (2012). *OSHA Trade News Release*, [online] Available at:<http://www.osha.gov/pls/oshaweb /owadisp.show_document?p_table=NEWS_RELEASES&p_id=21636>, [Accessed 13 March 2012].

Zeltzer, D., and Pioch, N.J. (1996). Validation, Verification of Virtual Environment Training Systems, *IEEE 1996 Virtual Reality Annual International Symposium*, March 30–April 3, 1996. Dayton, Ohio, pp.123–130.

Zhao,D, Thabet,W, McCoy,A and Klienr,B. (2012). Managing Electrocution Hazards in the US Construction Industry Using VR Simulation and Cloud Technology, *eWork and eBusiness in Architecture, Engineering and Construction*, ECPPM 2012, pp759-764.

Grand Challenges in Data and Information Visualization for the Architecture, Engineering, Construction and Facility Management Industry

Mani Golparvar-Fard[1], Pingbo Tang[2], Yong K. Cho[3], and Mohsin K. Siddiqui[4]

[1] Assistant Professor, Dept. of Civil and Environmental Eng., University of Illinois at Urbana-Champaign; PH (217) 300-5226; email: mgolpar@illinois.edu
[2] Assistant Professor, School of Sustainable Eng. and the Built Environment, Arizona State University; PH (480) 727-8105; email: tangpingbo@asu.edu
[3] Associate Professor, Durham School of Architectural Engineering and Construction, University of Nebraska- Lincoln, PH (402) 554-3277; email: ycho2@unl.edu
[4] Assistant Professor, King Fahd University of Petroleum and Minerals, Construction Eng Mgmt Dept, Saudi Arabia; PH +966 3 860 2325; email: mohsin@kfupm.edu.sa

ABSTRACT

This study aims to identify challenges in the AEC/FM industry that can potentially be addressed using research in data and information visualization. These challenges were recognized by reviewing visualization techniques for addressing current challenges associated with decision-making tasks; studying the fit between visualization techniques and the decision-making tasks; identifying the gaps in knowledge of visualization techniques; and finally establishing a framework for measuring the domain requirements and the technology capabilities as a road map for domain-needs-driven development in the area of data and information visualization. The challenges associated with the current practice of project delivery and the limitations of the state-of-the-art visualization techniques in addressing these challenges are discussed. This paper presents where and how intuitive and effective visualization can address these challenges. It also suggests areas where researchers can apply visualization research to further improve existing processes.

INTRODUCTION

In today's AEC/FM industry, project teams must collect and deal with a large variety of project data and information to support high-quality and timely decisions. Nonetheless, current representations do not effectively communicate project data and information, and have difficulties in highlighting their relationships. Consequently, project teams spend significant time mentally relating project data and information and manually performing decision analysis. The empirical observations on project teams using visualization techniques have anecdotally shown that the effective communication of project information and their relationship can significantly improve the decision-making process. There is a need to identify what kinds of visualization techniques are best suited for supporting various decision-making tasks in the delivery of AEC/FM projects. This study aims at identifying those project delivery challenges that can possibly be solved using data and information visualization tools.

RESEARCH METHODOLOGY

An extensive literature review was conducted to find those challenges faced by the practitioners within a typical project delivery lifecycle that can potentially be addressed through research in data / information visualization. These challenges and the limitations of the current visualization methods are presented in the following:

AEC/FM CHALLENGES FOR DATA/INFORMATION VISUALIZATION

During various stages of a project, practitioners are faced with the need to easily and quickly make decisions. These decisions are primarily based on the data and information that can be derived from the very large datasets required to represent the various facets of a project. How best data/information can be presented from these large datasets is a fundamental research question. This question is particularly important in the AEC/FM community, because data and information visualization tools are directly usable by practitioners for improving decision-making (Korde et al. 2005; Golparvar Fard et al. 2006; Chiu and Russell 2011). In the following, practical problems and the needs within each stage of a project are discussed:

1. Design Development, Design Coordination, and Value Engineering: The needs for visualization tools during programming, and schematic/detailed design phases can be divided into: 1) tools that facilitate collaboration in interactive workspaces wherein experts are working together in a physical space; and 2) virtual workspaces wherein experts are working remotely and functionalities can be synchronous/ asynchronous.

a. Physical Workspaces: Current challenges in physical workspaces include no direct cost feedback on design decisions, lack of coordination of trades, and inability for reusing design data for shop drawings because architects and structural engineers - due to different regulations - model buildings in different ways. There have been advances in detection tools that can find collisions between structural elements and MEP systems. However, without an as-built model of a physical space, it is still very challenging to apply these model checking tools to existing structures for remodeling or retrofit design purposes. Although laser scanning and photogrammetry have been used to create as-built models of existing facilities (e.g., Tang et al. 2012, Golparvar-Fard et al. 2011), their application is still labor intensive and the computational burden for processing data is still pretty high.

b. Synchronous/Asynchronous Virtual Workspaces: Virtual workspaces are cloud locations where data/information is shared through electronic means. Research has shown that these workspaces improve the utility of project information and the quality of the decision-making (e.g., Liston et al. 2001, Golparvar-Fard et al. 2006). Others have shown adoption of virtual tools such as Second Life can further enhance global virtual collaborations (Iorio et al. 2010). Current challenges include managing 3D BIM on interoperable protocols to support queries in multiple/ individual models; visualizing semantic models using web-based interfaces accessible in smart devices; enabling multiple-use real-time environments wherein users can combine partial BIMs, linking BIM data to GIS programs, and the ability to share with cloud-based structural analysis tools. Also there is no measurable account of usage patterns with qualitative measures of how visualization tools are used in productive ways (Iorio et al. 2010). The mismatch between the value of desktop sharing for design coordination in virtual spaces and the lack of use by design project networks are other challenges. Addressing these enables better understanding on how underlying functionalities built into virtual environments can be beneficial to virtual workspaces.

2. Model-based Cost Estimation: Visualizing buildings in 3D removes a good deal of uncertainty for cost estimation. It no longer takes an estimator with the knack of creating 3D buildings in his head based on 2D drawings to fully appreciate what s/he sees (Words & Images 2009). 3D Model-based cost estimation is proven to be a

leaner approach integrating BIM object attributes with cost databases (Tiwari et al. 2009). BIM ensures a unique source for cost estimating for the entire lifecycle (Azhar 2011). Compared with traditional approach, model-based cost estimation enables adjustment of the budget from one design scheme to another, estimates cost differences, and provides rapid feedbacks. Tiwari et al. (2009) points out that model-based estimating process will eventually be faster than traditional methods. The challenges of visualized cost estimating go beyond finding the right combination of software applications. Various types of cost databases are currently being used by the industry. Standards for the level of detail are not yet defined for many applications (Hartmann et al. 2008). Nowadays it is practically difficult to use same visual tools to estimate project cost for collaborative companies. Another challenge is the integration of workflows between 3D/4D and 5D applications (Forgues et al. 2012). The process of tallying 3D elements from spreadsheets and establishing relationships between quantity take-off items and recipes is time consuming and is prone to human error. Software interoperability might lead to such errors as well. How to use model-based cost estimation to generate visualizations for decision making processes is yet another challenge. Moreover most cost estimating tools do not support dynamic processes in early design phase where many changes are still possible (Tulke et al. 2008).

3. Scheduling and Constructability Reasoning: Despite the benefits of clash detection with 3D model and the detection of space-time conflicts using 4D models current clash prevention tools cannot easily visualize non-physical type of interferences (i.e., soft conflicts) concerning issues like clearances (e.g., cable tray requiring clearance on the top for access), or functionality (e.g., conduits blocking the air flow) which are critical in developing coordinated models. Current tools can identify conflicts that occur in clearance spaces around building elements (e.g., clearance spaces for access); however, to identify majority of the soft conflicts, more research needs to be done to explicitly consider the functionality, performance, tolerance, safety, and installation requirements (Golparvar-Fard et al. 2006). Recent studies by ENR (2012) shows the majority of the benefits observed from applications of BIM are still being reported in clash prevention. While 4D models can provide compelling visualizations for schedule quality control and contractor coordination, yet their applications are still limited. The main challenges are the lack of tools that can easily visualize the site layout beyond the actual construction related elements; the challenges associated with visualizing dynamic operations (e.g. equipment) or temporary objects (e.g. scaffoldings); and finally difficulties in quickly visualizing 4D models with the as-built information for facilitating contractor coordination processes.

4. Construction Performance Monitoring: a. *Progress/Quality Monitoring* require visualization of appearance-based and physical deviations between as-planned and as-built models. Research studies (e.g., Golparvar-Fard et al. 2011, Bosché 2010) have shown these models can be jointed visualized facilitating identification of deviations. Golparvar-Fard et al. (2009b and 2007) proposed application of metaphors based on traffic-light color to jointly visualize physical deviations and Earned Value Analysis metrics. In some cases, analyzing individual deviations and the correlations among them are important to to understand how progress/ quality issues are caused by shared reasons and how they influence one other. The challenges related to visualizing geometric differences lie in the computational complexity of detailed deviation

analysis. Recently developed spatial data structure and change analysis algorithms can efficiently handle detailed geometries captured in dense and large 3D data sets (e.g., point clouds generated by photogrammetric or laser scanning systems). Examples of such developments include Octree based change detection (Girardeau-Montaut et al. 2005; Zeibak and Filin 2007), feature-based spatial change analysis (Choi et al. 2009), and multi-resolution voxel for efficient neighborhood searching and comparison (Olsen et al. 2009; Truong-Hong et al. 2011; Golparvar-Fard et al. 2013). These efforts reduce the computational complexity of comparing spatial data sets capturing cm-level details, while some challenges of extra-large spatial data (e.g., Terabytes of 3D data sets collected by 3D mobile mapping systems) remain, including effective data decomposition, data interoperability, and large number of parameters that need to be configured by engineers (e.g., color map, neighborhood size) (Anil et al. 2012). Visualization technologies generate spatial distributions of deviations, but it is time-consuming and error-prone to have engineers manually configure geometric comparison algorithms and inspect deviations. Comparing non-graphic attributes between the as-planned/as-built objects requires reliable data-object association, nongraphic feature recognition (e.g. material) and semantic visualization. These challenges still involve some open questions. For congested sites with large number of elements, associating the as-built data points with the closest as-planned objects may be misleading (Bosché 2010; Tang and Akinci 2012a). Without reliable correspondences between data and model, the relevant nongraphic attributes of objects are not comparable. Research studies are addressing reliable classification of materials based on imagery data (Golparvar-Fard et al. 2012; Zhu and Brilakis 2010), while lighting conditions, material properties, and uncertain environmental conditions can still cause less reliable recognitions. Finally, visualizing differences may involve categorical attributes such as materials or strength levels which are not numerical and ordinal. Visualizing such differences is still difficult as methods for detecting scale and semantic issues of categorical differences are under development.

   b. Productivity and Operational Efficiency: Construction operations can be dynamically visualized by linking Discrete-Event Simulation and CAD models of the infrastructure, equipment, temporary structures, and other resources (Kamat et al. 2011). These models can serve as great communication tools for coordination, and validating the underlying logic of operational plans for maximizing productivity and operational efficiencies. Nonetheless, visualization of dynamic operations at a meaning-full level requires significant amount of operational details. Despite significant research in automated activity analysis over the past few years, more research needs to be done to enable the benefits of these visualizations.

   c. Environmental Impacts: Several new carbon footprint monitoring tools are recently developed (e.g., Shiftehfar et al. 2010; Memarzadeh et al. 2012) that enable practitioners to effectively benchmark, monitor, and visualize expected and released embodied/actual carbon footprint of a project. These visualizations provide practitioners with an opportunity during coordination, submittal, and onsite construction processes to not only insure timely delivery of materials, but also minimize the cradle-to-site and onsite carbon footprint of their projects. The main challenge associated with visualization for embodied carbon footprint lies in the need

for access to detailed datasets, and for actual carbon footprint lies in methods that can continuously track and localize equipment and their activities within 4D BIM.

d. Construction Safety: The challenges with visualization of construction safety in the design phase are related to challenges of modeling operational details and temporary structures such as scaffoldings and railings. During the construction phase, the challenges are rapid as-built modeling and integration with expected site models.

4. Building and Infrastructure Performance Metrics

a. Energy Performance Monitoring: Sustainable buildings have been a popular subject of study in the AEC/FM industry. The LEED rating requires energy modeling to assess the energy use of a building and to quantify the savings attributable to the proposed design. Currently an energy model can be directly created by importing BIM into a graphic energy modeling tool. For new buildings, BIM typically generated in the design phase can be directly used to conduct the energy performance analysis. Nonetheless, extensive manual modeling needs to be done in the design phase to import these models into the energy simulation tools. For accurate assessments, both geometry of the model and the semantics related to the thermal envelope (e.g., dimension and location of doors and windows, thermal zones, material types) have to be precise. For existing buildings, BIM is not always available. Even though BIM may exist, though these models might not be up-to-date. The preparation for new model is usually time-consuming, labor-intensive, and costly. Thermal zones and thermal information of the envelope also need to be manually obtained and assigned to the model for energy performance analysis purposes.

b. Health Monitoring: The challenges are mainly from the fact that large data sets are needed for capturing histories of monitored facilities, visualization of relative measurements, reliable health index generation and time series visualizations. Previous studies show the technical feasibilities of visualizing deformation histories of highway structures (Jaselskis et al. 2005), while pointing out the challenges of handling large data sets that are collected in multiple data collection sessions for capturing the life-cycle changes of such structures. For reducing the computational needs for visualizing deformation histories from large number of data sets, several studies use feature based approaches for monitoring purposes (Choi et al. 2009; Lim and Suter 2009; Teizer et al. 2005). These studies focus on visualizing geometric differences, while in many cases health monitoring agencies require tracking the changes of relative measurements (distance between certain features, alignments) (Tang and Akinci 2012b). Visualization of relative measurements still requires further explorations. Health monitoring requires visualizing the time series of health indexes of structures. Deriving structural index values from data is still an active research area (Olsen et al. 2009; Park et al. 2007). Furthermore, visualizing deterioration histories of structures require further explorations as to how to represent as-is histories together with the as-designed reference model. More research is needed to figure out how visualization of different parts of a structure with different health conditions and deterioration patterns can be handled.

5. Education in AEC/FM:
The need for collaboration in the AEC/FM industry is growing to meet the needs of rapidly delivering high quality projects with lower budgets (Salazar et al. 2006). The industry is currently facing training crisis with

aging workforce and little effort to utilize visualization techniques (Goedert et al. 2011). Gier (2007) believes that visualization tools within Construction Engineering and Management curriculum improve students' abilities on process. Salazar et al. (2006) state that the major benefits for students learning visualization tools include: 1) developing a better understanding of buildings and their components; 2) achieving advanced skill for industry career or research purpose. However, the primary challenge is that education has to face constant update to accommodate the emerging technologies, which could be heavy burdens to instructors. The selection of the software to be used for class should evolve appropriately with up-to-date standards (Sah and Cory 2009). Without prior experience with visualization tools, students might have low performance on those projects that benefit from visualization. Many researchers have pointed out that virtual learning environment through Internet are extended beyond physical reach of classrooms, and can complement traditional CEM education (Koskela et al. 2005; Sawhney and Mund 1998). Second Life platform and role-playing scenarios have been explored as means of communication tools to satisfy the interactive learning in classrooms. Goedert et al. (2011) developed a systematic, game-based framework for combining construction simulation scenarios and learning, but expanding the base and breadth of modulus for learning is pending to be addressed. While Second Life provides a communicating interface, it lacks functions for reviewing multiple designs from 3D models (Ku & Mahabaleshwarkar 2011)**.**

6. Frameworks for Assessment of Decision-Making Processes: Although most information is generated electronically for coordination meetings, teams still primarily communicate and share information using paper-based representations. Emerging technologies such as touch-sensitive large-screen displays, table-top displays, and tablet PCs offer great promise in enriching today's paper-based workspaces to create what are known as interactive workspaces (Golparvar-Fard et al. 2006). 3D BIM tools are also gaining acceptance and provide benefits to the coordination process. It remains unclear as to how such tools can be incorporated in workspaces to support coordination. There is still a need for frameworks that can assess the impact of visualization on decision-making and coordination meetings.

CHALLENGES IN VISUALIZATION TECHNIQUES

1. Data and Information Granularity/Quality: It is essential to define the correct level of detail by data range (e.g., entire vs. a subset) or granularity (e.g. all vs. spread footings). Not only range and granularity of data is important for effective visualization, but it is also a source of reference for resolving potential conflicts. Despite advancements in visualization, it is still challenging to efficiently process large amount of data and visualize the only required subsets.

2. Computation: The rapid improvements in computer software/hardware have provided engineers with powerful methods of storing, processing, sharing, retrieving and visualizing data (Abudayyeh et al. 2004). These improvements have made visualization as a growing and essential part in many areas, such as monitoring for safety, progress, productivity, and virtual walk through. Visualization technologies in areas such as point cloud/image/video processing, and augmented reality have also become prominent. In spite of the feasibility of visualization techniques, it is still challenging to efficiently process the massive data collected from the construction sites, which could include 3D point cloud, 2D images, video frames, and other sensor

data (Cho et al. 2012; Ham and Golparvar-Fard 2012; Gong et al. 2011). Point cloud data can easily reach hundreds of megabytes depending on the size of a construction site. The computation time exponentially increases when iterative algorithms are used to process the data. Another challenge is how to update visualization in real-time to assure the data authenticity for dynamic construction environments (Gai et al. 2012).

3. Interaction: It is essential to apply data fusion and conversion to satisfy user's needs. Wang et al. (2013) introduced a color coding method for fusing temperature data with point clouds. Golparvar-Fard et al. (2007) suggested overlaying as-planned models on photographs and formed a method for visualization of construction progress where deviations were represented in Augmented Reality environment such as ones developed by Behzadan et al. (2007 - 2009). Using metaphor of traffic light colors, progress deviations were visualized in comprehensive single imagery. The findings from these studies on as-built data collection, as-planned modeling and visualization of progress form a stepping stone upon which the proposed framework for interactive visualization of construction progress with the D4AR – 4D Augmented Reality- modeling method is developed. Such semantically-rich project models can support various project management and facility management functions. Integrating as-built and as-planned models also provides an opportunity to assess deviations between as-design and as-built conditions (Akinci et al. 2006).

CONCLUSION

This paper aimed at identifying those challenges across typical steps of a project delivery that can possibly be solved using data and information visualization tools. These challenges were discussed at individual stages within a project delivery and the gaps in knowledge from the perspective of data and information visualization tools were discussed. Due to page limit, complete evidence regarding the severity of the challenges is not presented. The different visualization methods discussed in this paper can be employed to make better decisions or to help solve current AEC/FM communication problems. The visualization challenges can also possibly help researchers define their research directions.

REFERENCES

Abudayyeh, O., Cai H., Fenves, S. J., Law,K., O'Neill, R., and Rasdorf, W. (2004). "Assessment of the Computing Component of Civil Engineering Education." *Journal of Computing in Civil Eng*, 18(3), 187-195.
Akinci, B., Boukamp, F., Gordon, C., Huber, Lyons and Park (2006). "A formalism for utilization of sensor systems & integrated project models for construction quality control." *Automation Construction*, 15, 124-138.
Anil, E. B., Tang, P., Akinci, B., and Huber, D. F. (2012). "Deviation analysis method for assessment of the quality of the as-is building information models generated from point cloud data." *J of Computing in Civil Eng.*
Azhar, S. (2011). "Building Information Modeling (BIM): Trends, Benefits, Risks, and Challenges for the AEC Industry." *Leadership and Management in Engineering*, 11(3), 241–252.
Bosché, F. (2010). "Automated recognition of 3D CAD model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction." *Advanced eng informatics,* 24(1), 107–118.
Chiu C. Y., and Russell A.D. (2011). "Design of a construction management data visualization environment: A top–down approach", *Automation in Construction*, 20 (4), 399-417.
Cho, Y. K., Wang, C., Tang, P., and Haas, C. T. (2012). "Target-Focused Local Workspace Modeling for Construction Automation Applications." *Journal of Computing in Civil Engineering*, 26(5), 661-670.
Choi, K., Lee, I., and Kim, S. (2009). "A feature based approach to automatic change detection from LiDAR data in Urban areas." *Laserscanning09*, XXXVIII, Paris, France, 259–264.
Engineering News Record (2012), *Business Value of BIM in North America*, McGraw-Hill Publications.
Forgues, D., Iordanova, I., Valdivesio, F. and Staub-French, S. (2012). "Rethinking the Cost Estimating Process through 5D BIM: a Case Study." *Construction Research Congress 2012,* ASCE
Gai, M., Cho, Y., and Wang, C. (2012). "Projection-Recognition-Projection (PRP) Method for Object Recognition from a 3D point cloud." 2012 ASCE Int. Workshop on Computing in Civil Eng, Clear Water, FL., 325-332.

Gier, D. (2007) *What Impact Does Using BIM Have on Teaching Estimating to Construction Mgmt Students*

Girardeau-Montaut, D., Roux, M., Marc, R., and Thibault, G. (2005). "Change detection on points cloud data acquired with a ground laser scanner." *Remote Sensing and Spatial Information Sciences*, 36 (3), W19.

Goedert, J., Cho, Y., Subramaniam, M., Guo, H., Xiao, L.(2011). "A framework for Virtual Interactive Construction Education (VICE)." *Automation in Construction* 20, 76–87

Golparvar Fard M., Sridharan A., Lee S. and Peña-Mora F. (2007). "Visual representation of visual representation of construction progress monitoring metrics on time-lapse photographs."

Golparvar Fard M., Staub-French S., Po B. and Tory M. (2006). "Requirements of a mobile interactive workspace for design development and communication." *Proc., ICCCBE*, Montreal, QC.

Golparvar-Fard, M., Peña-Mora, F., & Savarese (2011) "Integrated sequential as-built as-planned representation with D⁴AR tools in support of decision-enabling tasks" *J. of Construction Eng & Mgmt* 137(12), 1-21.

Gong, J., Caldas, C. H., and Gordon C. (2011). "Classifying Actions of Construction Workers and Equipment Using Bag-of-Video-Feature-Words and Network Models." *J. of Advanced Eng Informatics*, 25(4), 771-782.

Ham, Y. and Golparvar-Fard, M. (2012). "Rapid 3D Energy Performance Modeling of Existing Buildings using Thermal and Digital Imagery." *Proc., ASCE Construction Research Congress*, 991-1000.

Hartmann, T., Gao,., and Fischer (2008). "Areas of Application for 3D/4D Models." *J. Constr. Eng. Manage.*, 134(10), 776–785.

Iorio J., Peschiera G., Taylor J. E., & Korpela L. (2011). "Factors impacting usage patterns of collaborative tools designed to support global virtual design project networks", ITcon 16, 209-230.

Jaselskis, E., Gao, Z., and Walters, R. (2005). "Improving transportation projects using laser scanning." *Journal of Construction Engineering and Management*, 131(3), 377–384.

Korde, T., Wang, Y. and Russell, A. (2005). "Visualization of Construction Data", *Proc., 6th Construction Specialty Conference, Canadian Society of Civil Engineers*, Toronto, Canada.

Koskela, M., Kiltti, P., Vilpola, I., and Tervonen, J. (2005). "Suitability of a Virtual Learning Environment for Higher Education." *The Electronic Journal of e-Learning*, 3 (1), 21-30.

Ku, K. and Mahabaleshwarkar, P. 2011. "BIM for Construction Education." *ITcon*, Vol. 16 (2011), 189-208.

Lim, E. H., and Suter, D. (2009). "3D terrestrial LIDAR classifications with super-voxels and multi-scale Conditional Random Fields." *Computer-Aided Design*, 41(10), 701–710.

Liston, K., Fischer, M. and Winograd, T. (2001). "Focused Sharing of Information for Multidisciplinary Decision Making by Project Teams", *Journal of ITCON*, 6, 69-82.

Memarzadeh, M., & Golparvar-Fard, M. (2012). "Monitoring and Visualization of Building Construction Embodied Carbon Footprint Using DnAR Models" *Construction Research Congress*, 1330-1339.

Olsen, M. J., Kuester, F., Chang, B. J., and Hutchinson, T. C. (2009). "Terrestrial Laser Scanning-Based Structural Damage Assessment." *Journal of Computing in Civil Engineering*, 24(3), 264–272.

Park, H. S., Lee, H. M., Adeli, H., and Lee, I. (2007). "A New Approach for Health Monitoring of Structures: Terrestrial Laser Scanning." *Computer-Aided Civil and Infrastructure Eng*, 22(1), 19–30.

Sah, V. and Cory, C. (2008) "Building Information Modeling: An Academic Perspective", *IAJC-IJME*

Salazar, G., Mokbel, H., and Aboulezz, M. (2006). "The building information model in the civil and environmental engineering education at WPI." *Proc., Eng Education and Practice for the Global Community*,.

Sawhney & Mund (1998) "Simulation Construction Management Learning Systems" *WinSim*,1319– 1324.

Shiftehfar, R., Golparvar-Fard, M., Peña-Mora, F., Karahalios, K. G., & Aziz, Z. (2010). "The Application of Visualization for Construction Emission Monitoring". *Construction Research Congress,* 1396-1405.

Tang, P., and Akinci, B. (2012a). "Automatic execution of workflows on laser-scanned data for extracting bridge surveying goals." *Advanced Engineering Informatics*.

Tang, P., and Akinci, B. (2012b). "Formalization of workflows for extracting bridge surveying goals from laser-scanned data." *Automation in Construction*, 22(3), 306–319.

Teizer, J., Kim, C., Haas, C., Liapi, K., and Caldas, C. (2005). "Framework for Real-Time Three- Dimensional Modeling of Infrastructure." *Journal of the Transportation Research Board*, 1913(-1), 177–186.

Tiwari, S., Odelson, J., Watt, A., & Khanzode, A. (2009). Model Based Estimating to Inform Target Value Design *<http://www.aecbytes.com/buildingthefuture/2009/ModelBasedEstimating.html> (Dec.21, 2012)*

Truong-Hong, L., Laefer, D. F., Hinks, T., & Carr, H. (2011). "Flying voxel method with Delaunay triangulation criterion for façade/feature detection for computation", *Journal of Computing in Civil Engineering*, 1, 132.

Tulke, J., Nour, M. and Beucke K. (2008) "A Dynamic Framework for Construction Scheduling based on BIM using IFC", *Proc. 17th Congress of IABSE*, Sept. 17-19, Chicago, USA

Wang, C., Cho, Y., and Gai, M. (2013). "As-is 3D Thermal Modeling for Existing Building Envelopes Using A Hybrid LIDAR System." *Journal of Computing in Civil Eng*, (doi: 10.1061/(ASCE)CP.1943-5487.0000273)

Zeibak, R., & Filin, S. (2007). Change detection via terrestrial laser scanning. *International Archives of Photogrammetry and Remote Sensing*, 36(3/W52), 430-435.

Zhu, Z., and Brilakis, I. (2010). "Concrete Column Recognition in Images and Videos." *Journal of Computing in Civil Engineering*, ASCE, 24(6), 478–487.

Visualizing Bridge Inspection with 2D+1 Software

Richard E. Lindenberg, S.E.[1] and Jonathan C. McGormley, P.E., S.E.[2]

[1]Senior Associate, Wiss, Janney, Elstner Associates, Inc., 330 Pfingsten Road, Northbrook, IL 60062; phone (847)272-7400; email: rlindenberg@wje.com
[2]Principal, Wiss, Janney, Elstner Associates, Inc., 330 Pfingsten Road, Northbrook, IL 60062; phone (847)272-7400; email: jmcgormley@wje.com

ABSTRACT

A true challenge in the digital age is the fusion of qualitative and quantitative information for meaningful visualizations and interpretations. Nowhere is this more obvious than in the inspection of bridge structures, where multi-piece data is collected for repetitive structural elements with the purpose of differentiating data content for condition assessment. Inspection by experienced personnel remains a critical and necessary time-consuming element. A true 4D BIM approach remains a challenge in the field due to hardware requirements and user skill demands.

A new 2D+1 approach has been implemented with a software system on lightweight digital tablets designed to annotate plan documents and photos. The software captures quantitative and qualitative observations (e.g. photos and written observations) referenced spatially to 2D drawing/photos of the structure with near real-time multi-user synchronization. A cloud-based system provides a near real-time replica of the document to most accessible web browsers. Quantitative, as well as qualitative data can be analyzed programmatically together using spreadsheet software or developed into more advanced visualizations through database querying techniques. The software already has been used to improve the efficiency and quality of data collection on numerous bridge inspections across the U.S.

BACKGROUND

The National Bridge Inspection Program has over 600,000 bridges and culverts in the US with spans exceeding 20 ft. (USDOT, 2010), which require visual inspection at least once every twenty-four months. Many of these structures are 50 years and older (National Bridge Inventory, 2011). The visual inspection process consists of trained inspectors collecting multi-piece data for repetitive structural elements with the purpose of differentiating data content for condition assessment. Recent changes to inspection protocols through MAP-21 (FHWA, 2012) expands further the type and quantity of data collected through element level inspections to aid in data-driven, risk-based decision making. Inspectors traditionally use paper (i.e. drawings of the structure and field notebooks), pencil, and a camera.

On the other end of the spectrum, new bridges are being designed and constructed with state-of-the-art Building Information Management (BIM) Software, which is

capable of housing everything from the exact three-dimensional (3D) geometry to wattage requirements of the channel lights. Some have advocated trying to collect existing bridge information and model it inside of BIM systems (DiBernadardo, 2012). There would be a great deal of value in collecting this information; however, the added cost to input this information would arguably burden already stretched state DOT budgets. In addition, it would be unreasonable to expect the inspectors to input it in the field given the software expertise required and the current hardware requirements (Autodesk, 2013).

A new idea has been developed to improve inspection efficiency, provide better quality assurance, and allow more sophisticated analysis using the concept of 2D+1 software. BIM is traditionally thought of as 4D software (GSA, 2007), because in addition to housing the 3D geometry it also acts as a database of relevant document information over time. Thus, information can be spatially organized, and a variety of professionals can utilize and interact with the information database.

The concept of 2D+1 is much the same, except instead of 3D model geometry, existing two-dimensional (2D) documents such as plan drawings or photos are used. The 2D documents serve as the spatial reference for software that provides quick tools for the bridge inspector to add relevant information. The collected data is exported in an Extensible Markup Language (XML) format that allows analysis tools to utilize all the aspects of the information, including the inspector's comments, photographs, and the spatial reference point of the observation. This provides almost unlimited data density of the digital document, such that many details and historical inspection information can be included in the documents. The 2D+1 prototype software and the particular features discussed herein has been developed by Wiss, Janney, Elstner Associates, Inc. and implemented in the field on several large interstate bridge inspection projects.

SOFTWARE AND HARDWARE

The software and hardware for the inspection system focus on keeping it lightweight for the inspector who has to use it all day and carry it in an unforgiving environment. The 2D+1 software system is comprised of two major components that optimize the field inspection process. The 2D+1 data collection in the field is performed using a custom developed Apple iOS application for the Apple iPad (iPad). Selection of the hardware was based on the long battery life, relatively low-cost, and durability when properly protected. To date, the hardware has demonstrated it's performance and reliability in the bridge environment with the authors supervising over several thousand hours of bridge inspections. In addition, the software can upload photos to a central storage location and provide references in the digital documents utilizing many standard digital cameras or the iPad's internal camera. Final photo and document storage, as well as a digital document viewer for the internet, is provided through an internet accessible web server. These internet services comprise the cloud computing capabilities. While the cloud server is not required for the inspection, it provides document and user administration, as well as user or version synchronization.

FREEFORM INSPECTION

The concept of digital data collection, even for inspection, is not particularly novel. Computer database software such as Filemaker (2013) readily allow a developer to develop a form and database to collect prescribed information and even do it on a mobile device such as an iPad. In addition, software developers have also come up with proprietary software to assist with the bridge construction inspection process, such as tracking requests for information. These software tools have a narrow focus, and are not designed to adapt to unknown inspection conditions.

OPTIMIZED FIELD DATA COLLECTION

The 2D+1 software works by taking a PDF drawing or photo (e.g. plan, elevation, or detail) and allowing the inspector to annotate it with a variety of tools varying in complexity from simple primitive graphics to more advanced line drawing tools (i.e. polygonal and freehand). The 2D+1 software allows for customizable data entry forms—a capability that separates this software from most others. The inspector can select amongst pre-defined inspection forms established to meet DOT-specified inputs to more open-ended comment type inspection forms. Each form provides the ability to add multiple photos taken by the inspector. Prescribed inspection point information can further be optimized with drop-down selection tools to save data entry time.

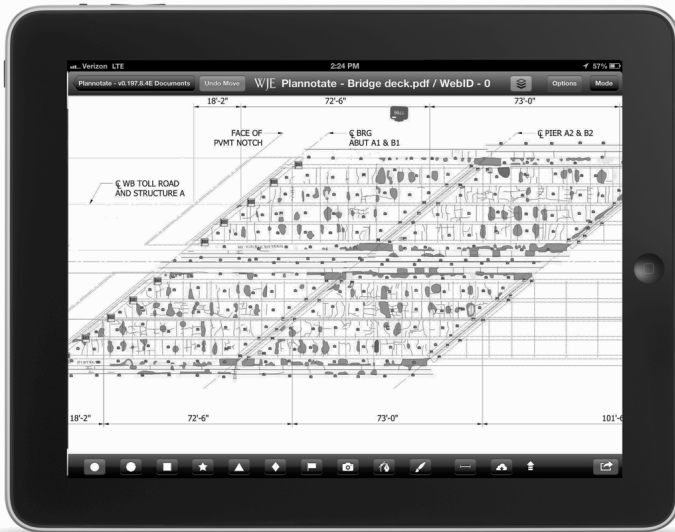**Figure 1. View of a bridge deck with deterioration annotated with the 2D+1 software.**

Figure 1 shows an example of a bridge deck survey with hundreds of freehand drawn area annotations representing concrete delaminations and spalls, as well as lines representing cracking. Both the top surface and underside survey information can be contained in the same document, and collected at the same time in the field. Thus,

inspectors often can see correlations of the deterioration soon after the observations are made. Another critical aspect is that deterioration photos are not attached simply as photo objects, but rather attached to the relevant annotations representing the observations. For example, a photo of a concrete spall would be attached to the freehand drawn annotation that approximates the size and location of the spall on the plan. Thus, there is a fusion of the actual visual image of the deterioration to a spatially-located image. The software can even calculate the size of the inspectors freehand drawn spall annotation based on a scale set in the document.

Likewise as shown in Figure 2, digital photos themselves can also be annotated as documents to show more in-depth information. Annotations provide exactly the same form information collection capability, allowing the inspector to provide information about the area being marked. Furthermore, these photo documents can be linked to a general plan overview document to create a nest or network of documents. This allows for a great deal of complexity, as well as organization to the inspection.

Figure 2. Digital photo taken in the field with a freehand drawn annotation, which shows deteriorated area of a cover plate.

Time spent in the field carrying out bridge inspections is influenced by traffic control requirements and the desire to limit impact to the traveling public. The prototype software has been revised over the past year based on feedback from hundreds of hours of field use to optimize the speed and ease of use. Through generational improvements, inspectors can now complete the field portion of the bridge inspection with similar time requirements as a traditional inspection. However, the 2D+1 data represents an enormous improvement in the overall process since it greatly reduces the time required in the office over the typical data reduction process. Furthermore, the system's cloud server provides the inspector with near real-time quality control of

data collection, thus reducing the potential of missing information. Multiple inspectors can work more efficiently through better coordination. In addition, the inspection program manager and even the bridge owner can remotely interact with the digital version of the collected data as the inspection progresses. As inspectors identify issues requiring special attention or clarification, they can contact technical experts and provide those experts with up-to-date and organized documents with digital images. The expert can participate from anywhere in the world with an internet connected web browser.

STRUCTURED DATA AND ANALYSIS

The core concept behind the 2D+1 software is that the data comprise a database with spatial references to the relevant base documents. So instead of developing pre-defined reports, the focus is on developing a useful data structure. The prototype software uses an XML structure to warehouse and manage the relationships of the data collected. The XML format provides many relational database benefits; however, the data structure is able to grow dynamically without the constraints of traditional structural relational database constraints.

Since the data structure forms a database, the data is capable of being queried in a number of software programs. For example as shown in Figure 3, data can natively be imported into Microsoft Excel using the built-in XML import. The graphical and spatially referenced data is then transformed into a tabular format. Photos are referenced to the annotations and the XML structure preserves these relationships. The result is that the photo maintains a critical link to the inspector's observation, and not merely a photo at a spatial referenced point. So when working with tabular data, the sorting of concrete deterioration, for example, can be reviewed from largest to smallest quantity. In turn, the engineer can focus on the areas of greatest interest while accessing the associated photo(s) in the prioritized list.

The XML structure also provides the ability to use XML database tools, such as BaseX (2013) to aggregate data within a single document or across thousands of bridge inspection documents. The BaseX software provides access to powerful XML querying tools such as XPATH or XQUERY. In addition, it also provides an application programming interface (API) to allow programmatic access. As an example in Figure 4, annotations from over twenty documents from a bridge inspection are searched and the locations are graphically represented. The filtered and aggregated data can then been organized and outputted. Now prior to an inspection team leaving a bridge site, all documents can be aggregated and searched to assure that all areas have been covered, as well as to provide preliminary quantity and condition state information. The aggregation of inspection information prior to leaving a bridge allows the owner and program manager to better adapt the inspection process to the investigation findings.
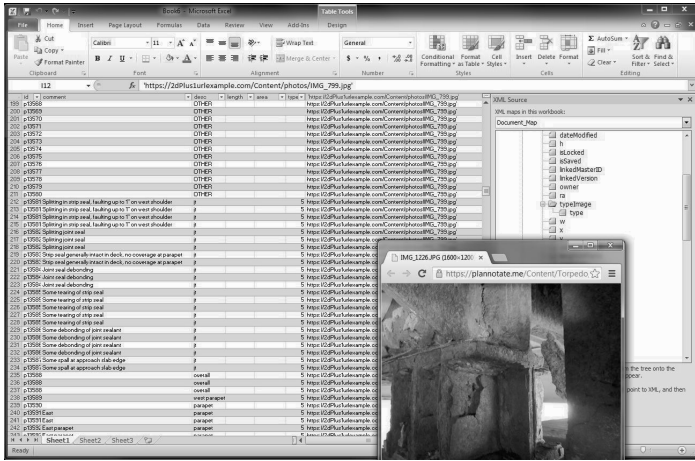
Figure 3. Direct XML import into Excel for data analysis.

DATA VISUALIZATIONS

Since all of the observation data are tied spatially to the document and stored in a readable XML format with coordinate information, a number of powerful analyses options become available. Annotations store multi-piece information for each inspection point, including measurements, descriptions, and answers to prescribed input. This can create an enormous data density within the documents. However, when it comes time to understand and evaluate information, the engineer needs to filter the information and look at the individual performance of single variables. Here the 2D+1 data concept comes to bear, through simple query techniques, annotations can be transformed. Utilizing Mathematica, authors have developed numerous visualizations by transforming the annotations into spatially overlaid information content. The same document in Figure 1 could have the values of the measured areas of deterioration placed at the centroid of the drawn annotations. The engineer and the owner now have the deterioration information quickly overlaid onto a relevant document that is easy to understand. Figure 5 shows a segment of several miles of an elevated bridge structure where a single observation annotation was placed in each span. Deterioration quantities were then overlaid using the location of each of the spans. In addition, the prototype software can provide references to other documents, so repair drawing documents could also be linked. This provides quick and organized access to many documents.

Far more advanced analysis can be leveraged by combining the spatial information with traditional deterioration measurements. For example, a drawn boundary can be placed to mark out areas of concern—the estimated negative bending moment region of the bridge's deck. An analysis can then be run using the software to identify all annotations within the boundary. These results could then be aggregated and

visualized for reporting. Historical information can also be readily queried to evaluate the advancement of deterioration.
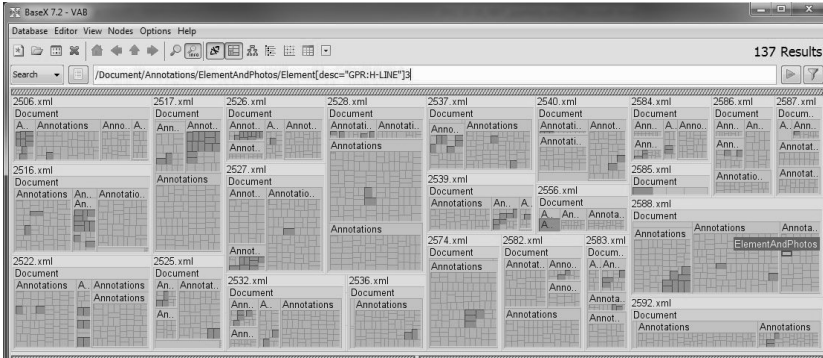
Figure 4. A summation query across thirty documents in a BaseX XML database for NDE annotations.

The benefits of the digitized inspection go beyond the inspection with the ability to also transform the inspector's data for subsequent repair documents. In the prototype system, a software algorithm was written to import annotations into Autodesk AutoCAD organizing the various deterioration information and observations by layers. This provides the ability to improve the follow-up repair process by developing repair documents based directly on the inspector's written observations. Furthermore, the drawings of record can be maintained by the DOT's in their traditional systems.
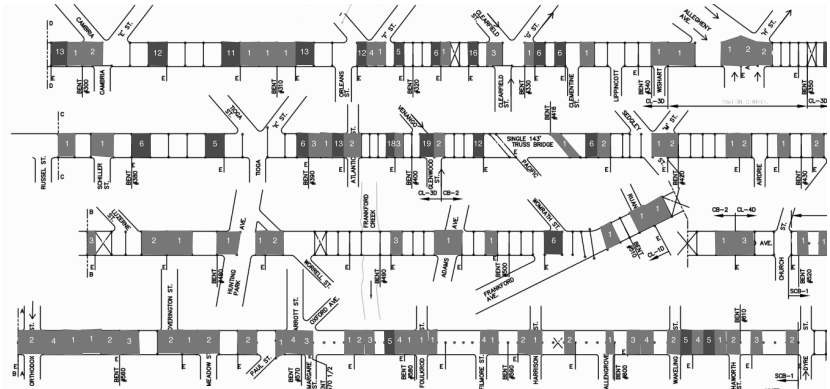
Figure 5. Elevated bridge structure with deterioration data dynamically mapped using Mathematica.

CONCLUSION

A lightweight tablet solution using a 2D+1 approach for marking up existing plans using overlaid annotations with underlying data structure has proven to be efficient and economical in bridge inspection with real benefits throughout the process. This especially holds true for the relative ease of training inspectors, a free-form data entry approach, and the ability to have real-time access for improved quality assurance. Back in the office, the software digitization improves the ability to better understand the current bridge condition state through analysis and data visualization techniques, as well as to develop repair documents.

REFERENCES

Autodesk. (2013). "System Requirements for Autodesk Revit Products."
http://usa.autodesk.com/adsk/servlet/ps/dl/item?siteID=123112&id=15385625&linkID=9243099 (accessed January 19, 2013).

BaseX. (2013). "BaseX. The XML Database." http://basex.org/home/ (accessed January 2, 2013).

DiBernadardo, S. (2012). "Integrated Modeling Systems for Bridge Asset Management." ASCE 2012 Structures Congress, International Bridges Session, Chicago, IL, March 31.

Filemaker. (2013). "Filemaker: Complete Platform to Create, Deploy, and Manage Custom iPad and iPhone Solutions for Business."
http://www.filemaker.com/products/filemaker-go/docs/fm_ios_complete_plat_wp_final.pdf (access Jaunary 2, 2013).

GSA. (2007). "BIM Guide Overview."
http://www.gsa.gov/graphics/pbs/GSA_BIM_Guide_v0_60_Series01_Overview_05_14_07.pdf (access January 2, 2013)

National Bridge Inventory (2011). http://www.fhwa.dot.gov/bridge/nbi/yrblt11.cfm (access January 19,2013)

USDOT (2010). "2010 Status of the Nation's Highways, Bridges, and Transit: Conditions & Performance." Report to Congress, Federal Highway Administrations, Federal Transit Administration, U.S. Department of Transportation, http://www.fhwa.dot.gov/policy/2010cpr/ (access January 19, 2013)

FHWA (2012). Moving Ahead for Progress in the 21st Century Act (MAP-21). Federal Highway Administration
Office of Policy and Governmental Affairs
July 17, 2012

Target-free Automatic Point Clouds Registration Using 2D images

Mengmeng Gai[1]; Yong K. Cho A.M.ASCE[2]; Qinghua Xu[3]

[1]GRA, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska - Lincoln, 1110 S. 67th St. Peter Kiewit Institute (PKI) 118A, Omaha, NE 68182, USA; PH +1-402-206-7617; FAX +1-402-554-3850; Email: mengmeng.gai@gmail.com;
[2]Associate Professor, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska-Lincoln, 1110 S. 67th St. Peter Kiewit Institute (PKI) 104C, Omaha, NE 68182, USA; PH +1-402-554-3277; FAX +1-402-554-3850; Email: ycho2@unl.edu; (Corresponding author)
[3]GRA, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska - Lincoln, 1110 S. 67th St. Peter Kiewit Institute (PKI) 118A, Omaha, NE 68182, USA; PH +1-718-664-8432; FAX +1-402-554-3850; Email: qinghua.xu@huskers.unl.edu

ABSTRACT

This paper presents a target-free automatic point cloud registration method which is based on identified common features between multiple images in point clouds. Target-free automatic registration of point clouds is an important research area in which several problems are still not addressed completely, such as registration accuracy, time, and size of overlapping area between point clouds. In this study, a series of point cloud data sets from different scan positions are registered based on the texture data in the overlapping area between point clouds sets. Both texture and point cloud data are obtained by a developed Laser Detection and Ranging (LADAR) system simultaneously. The automatic registration algorithm begins by identifying 2D common features of the texture, then constructing corresponding 3D common points, and afterwards finding out the optimal transformation matrix. The introduced method will reduce substantial amount of time for point clouds registration by using a few images within a small overlapping area between point clouds without the need of target references, thus significantly promoting as-built modeling process for existing infrastructure or construction progresses.

INTRODUCTION

Rapidly updating as-built design with detailed 3D scenario is useful in construction applications such as process monitoring and safety hazards detection

(Huber et al. 2010, Teizer et al. 2010, Golparvar-Fard et al. 2011). 3D laser scanning technology has been intensively used to render real-size objects or environment by registering point cloud data onto the same coordinates.

Point cloud registration is defined as registering multiple point clouds scanned from different viewpoints into one common coordinate system. The current state-of-the-art approach is to find at least three common points between two overlapped point clouds, and then calculate 3D rigid transformation matrix based on these three common points. Many types of commercial software are now available to realize the registration function by manually assigning three common points. However, this manual process is time-consuming and inaccurate when the data sets are huge and complicated. In order to tackle the registration problem, various methods have been proposed. The most popular idea is to manually or automatically find at least three common points using targets between two overlapped point sets, and then calculate the 3D rigid transformation matrix based on these common points. The manual processes of placing, maintaining, and relocating targets are time-consuming and could inevitably cause errors, which affects the quality of registration results. As for automatic registration, due to large size of point cloud data sets and limited processing rate, the registration process is normally time-consuming, which inevitably reduces updating speed of 3D points. Many prevalent ideas have been testified by researchers to improve automatic registration algorithms. The traditional Interactive Closest Point (ICP) algorithm (Besl and Mckay 1992) uses point-to-point correspondences to calculate closest distance of two point sets until it reaches certain criterion. Obviously, ICP-based approach is time-consuming when dealing with large data sets. Li and Wang (2008) introduced weighted value to the process of finding relative control points to improve the accuracy. Geometric Primitive ICP with the RANSAC (GP-ICPR) was proposed by (Bae and Lichti 2008) using the change of geometric curvature, approximate normal vector of the surface, and neighborhood search to improve the efficiency and accuracy. Parallel computing of MapReduce was presented by (Liu and Xie 2011) to improve the efficiency of computing. Men et al. (2011) integrated values with ICP algorithm to develop a 4D ICP algorithm, in which the Hue value was calculated according to RGB data captured by a digital camera. With the assistance of the Hue value, the ICP algorithm can be improved by getting higher accuracy and faster convergence. The accuracy of ICP algorithm depends on data quality and the size of the overlapping areas and requires the initial position of two point sets to be close enough, which is difficult to achieve in some circumstances.

Advancing feature-based registration can be realized without knowing initial starting points, and utilize the 2D image processing technology to assist the recognition of feature points. However, many more scans are needed to achieve better performance. Encourage by Scale-invariant feature transform (SIFT) algorithm, Bay et al. (2006) presented Speeded-Up Robust Features (SURF) method using detection, description, matching scheme. This innovation algorithm based on integral image reduces feature computation and matching time. Eo et al. (2012) also utilized the feature points extracted from 2D intensity images using SIFT algorithm. However, this method still depends on the size of overlapping area. In their test, 12 scans were collected for registering one corner of the building. The accuracy of transformation matrix was within 0.005–0.069 m. The heavy computation load is another limitation

for feature-based registration. Thousands of feature points can be extracted from each scan based on geometry or image information, while most of them will not be filtered out because of the wrong or low accuracy match.

The main objective of this paper is to develop a target-free automatic point cloud registration method, based on identified common features between multiple images in point clouds. A feature-based algorithm was provided in this study to increase registration speed and accuracy, which automatically matching 3D point cloud data by utilizing 2D common features. The input data were collected from a 3D laser scanner equipped with a digital camera. Both devices are mounted on a same platform and rotate 360 degree at the same time, so the point could data and digital images at certain angles are correspondent. The camera was calibrated to avoid distortion in image processing stage. The intrinsic matrix of the calibrated camera will be applied when calculating fundamental transformation matrix.

The method structure will be introduced firstly, and followed by the process of points and texture data fusion. Then, the point cloud registration method will be discussed. Finally, conclusions and future work will be given.

METHODS

This study proposes a new feature-based automatic data registration framework to minimize overlapping areas and accurately and rapidly estimate common feature points from the overlapping areas from two scan sets. In this research, a new target-free automatic registration method was developed by utilizing the Speeded-Up Robust Features (SURF) algorithm and Kinematics theory. The overall framework for automatic registration is shown in Figure 1.
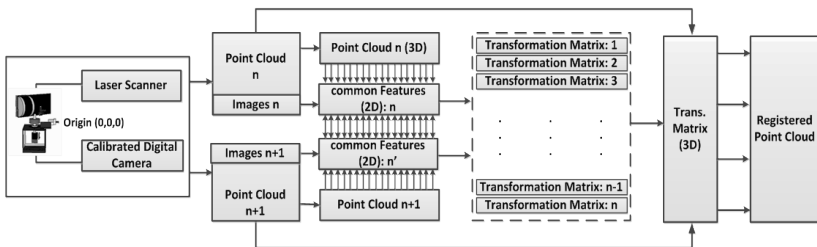
Figure. 1 Proposed automatic registration framework

A LADAR system provides point clouds of surroundings, together with texture data (images) captured by a calibrated digital camera. Point clouds and texture data share a same base-coordinate, located in the axle center of the step motor. Texture data are mapped onto the surface of corresponding point clouds according to the kinematical relationship. SURF descriptor is employed to extract common features among a series of texture data. Texture data from overlapped area produce more common features than other area. Common features extracted from 2D texture data are corresponding to 3D common points from point clouds, which are filtered

through developed triangle-based method by removing outliers. Three common points are randomly selected from each point cloud data sets to calculate the transformation matrix. The optimized matrix will be selected from the itinerary process to register point cloud data.

Points and Texture Data Fusion

Video camera captures texture data of surroundings, which are mapped onto the surfaces of the 3D point clouds. In the texture mapping process, each point is considered as an object containing different types of data, such as x-y-z coordinates. First, the concept of perspective projection was introduced, and a test was given to create a reference plane. During the reference test, a 120cm×90cm rectangular object was used, and both the laser rangefinder and the video camera were placed parallel to the object. The distance between object and system was adjusted to make sure that the object, like a hydraulic excavator, completely filled the video camera's view. Assume the distance between object and camera is f_{ref}; the view angle of the video camera is θ, and (X_i, Y_i, Z_i) represents the coordination of the point cloud in the system coordinate system. Having the reference plane, all the objects parallel to the system could be correctly mapped according to Equation 1:

$$X_{pix} = \frac{X_i - \left(X_c - \frac{120 \times |Z_i|}{2 \times f_{ref}}\right)}{Pixel\ Unit \times \frac{|Z_i|}{f_{ref}}}, \qquad Y_{pix} = \frac{Y_i - \left(Y_c - \frac{120 \times |Z_i|}{2 \times f_{ref}}\right)}{Pixel\ Unit \times \frac{|Z_i|}{f_{ref}}} \qquad \text{(Equation 1)}$$

where X_c and Y_c are the coordinates of the point obtained when the laser is in its default position. X_{pix} and Y_{pix} are the coordinates of a pixel in the 2D digital image. Once the coordinates of the corresponding pixel were found, the pixel and its reference X and Y position in digital images were projected to the point in the 3D point cloud.

The video camera pans and tilts to obtain pixel data of another part of the point cloud. Under this circumstance, the reference plane is no longer parallel to the object. Due to the effect of the perspective projection, objects in the distance appear smaller than objects close by. If a simple interpolation was used and steps were equally spaced to compute pixel coordinates, a distorted image map would result. To avoid such a problem, the perspective correction method was used in this research. Perspective correction mapping interpolates after dividing by depth Z_i, then uses the interpolated reciprocal to recover the correct coordinate (Hill and Kelley 2006):

$$X_\alpha = \frac{(1-\alpha)\frac{X_a}{Z_a} + \alpha\frac{X_b}{Z_b}}{(1-\alpha)\frac{1}{Z_a} + \alpha\frac{1}{Z_b}}, \quad where\ 0 \leq \alpha \leq 1 \qquad \text{(Equation 2)}$$

When the video camera is rotated, the area of the camera view will vary based on the angle at which the camera is rotated. The coordinates of the edge points in the digital image can be calculated using Equation 3:

$$X_a = X_c + Z_c \tan\left(\alpha - \frac{\theta}{2}\right), X_b = X_c + Z_c \tan\left(\alpha + \frac{\theta}{2}\right) \qquad \text{(Equation 3)}$$

When the objects are not parallel to the camera, therefore, the pixel value with reference positions can be mapped to the 3D point cloud using the following equation:

$$X_{pix} = \frac{X_i - X_a}{Pixel\ Unit \times \frac{|z_i|}{f_{ref}} \times \frac{1}{X_\alpha}}, Y_{pix} = \frac{Y_i - Y_a}{Pixel\ Unit \times \frac{|z_i|}{f_{ref}} \times \frac{1}{X_\alpha}} \qquad \text{(Equation 4)}$$

During the scan, each point collected by the LADAR was considered as an object to find the corresponding pixel value according to the equations above.

Point Cloud Registration

In order to find at least three common points between two overlapped point sets, and calculate the 3D rigid transformation matrix from common points, SURF descriptor was employed and implemented in the overlapped point cloud area. Figure 2 shows an example of extracted common features from two images using SURF descriptor.

To improve the accuracy, various methodologies with promising results were proposed to filter the outliers, such as the Random Sample Consensus (RANSAC) algorithm (Fischler and Bolles 1981). Based on the RANSAC algorithm, there have been a number of researchers who were trying to improve RANSAC's verification and sampling performance, such as Matas and Chum (2004) and Capel (2005). These efforts aimed to optimize the model verification and sampling process to generate much more meaningful hypotheses. In particular, Nister (2003) proposed a system aiming to perform robust estimation and find the best solution from a set of hypothesis, together with a preemptive scoring based on an inlier-outlier model. However, inherent non-adaptive performance on the data array becomes a limitation and for low contamination problem, this framework seems slower than the standard RANSAR (Rahul, et al 2008).

In this study, a triangle relationship based filtering method was developed to remove the outliers from the extracted common features. The main goal of registration is to calculate the 3D transformation matrix between two groups of points. The calculation process is based on the filtered triangles produced in the above phases. Take one group of triangle A1A2A3 and B1B2B3 as an example, first, set point A1 as (0,0,0), move point B1 to the point A1. The final transformation matrix can be summarized as

$$T = \begin{bmatrix} T_1 & T_2 & T_3 & Dx \\ T_4 & T_5 & T_6 & Dy \\ T_7 & T_8 & T_9 & Dz \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{(Equation 5)}$$

Where matrix T is the transformation rotation item and D is the offset between two sets of point clouds. Figure 2 (c) and (d) show an example of registration results between two sets of point cloud data. The registration resolution is listed in Table 1. The registration speed is mainly determined by the common features extraction process, and the point cloud registration is implemented based on the transformation matrix generated from the common features. As a result, the whole registration process takes second level of time.

Table 1. Registration Resolution Results

	X Axis	Y Axis	Z Axis
Distance offset(mm)	12.16	0.37	0.50
Angle error(degree)	1.13	0.03	0.95

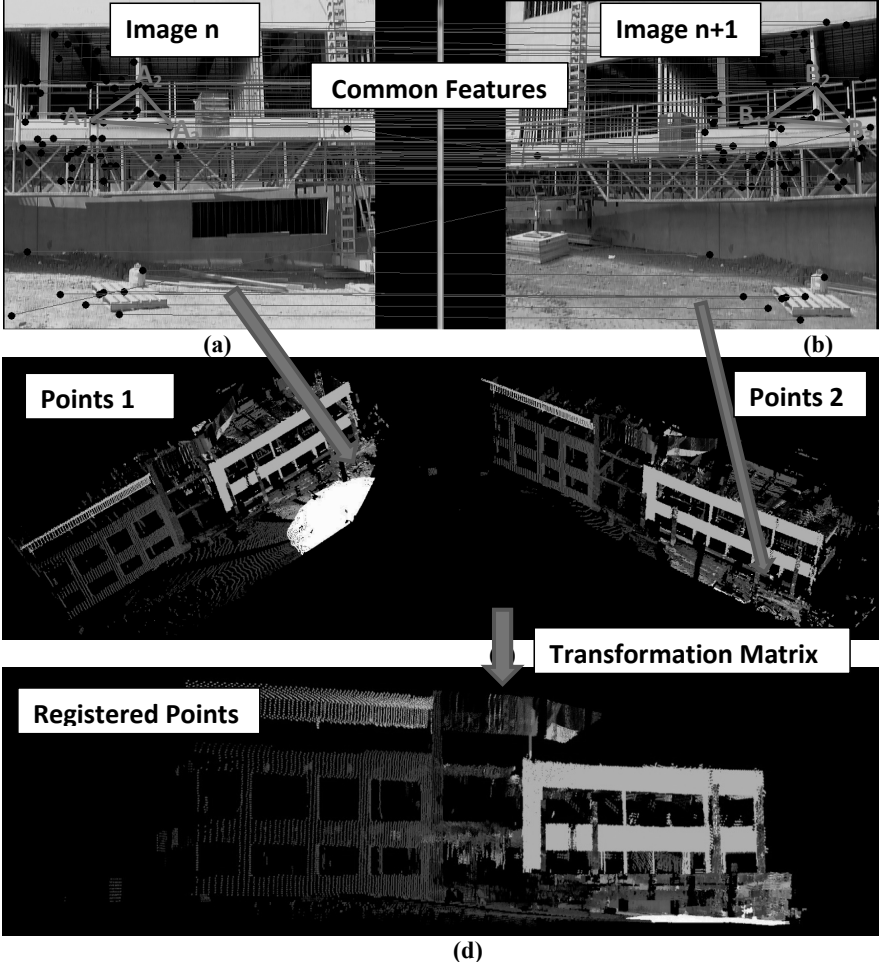(a)   (b)

(d)

Figure 2. Point Cloud Registration Process and Results

CONCLUSIONS

A novel target-free point cloud registration method was developed using SURF descriptor and triangle-based outlier filtering method was introduced in this paper. A LADAR system equipped with a digital camera was utilized to obtain point clouds together with mapped texture data. Common features were extracted from texture data in the overlapped area and used to calculate the transformation matrix. Point clouds were registered based on the matrix. The goal of this method is to automatically register two point cloud data sets with small overlapping areas, reduce computational time to seconds, and increase accuracy and efficiency of registration process. The main contribution of this paper is to introduce a common feature based point cloud registration method. To evaluate the validity of registration results, the distance error ranges from 0.37mm to 12.16mm, and the angle error ranges from 0.03 degrees to 1.13 degrees.

ACKNOWLEDGEMENT

REFERENCES

Bay, H., Tuytelaars, T., Gool, L.V. (2006). "SURF: Speeded up robust features." *Computer Vision–ECCV*, 404-417.

Bae, K. and Lichti, D. D. (2008). "A method for automated registration of unorganised point clouds." *Journal of Photogrammetry & Remote Sensing,* ISPRS, 63 (1), 36-54.

Besl, P. J., and McKay, N. D. (1992) "A method for registration of 3-D shapes." *Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 14(2), 239–256.

Capel, D.(2005). "An Effective Bail-Out Test for RANSAC Consensus Scoring." *Proc. British Machine Vision Conf*, pp. 629- 638.

Eo, Y. D., Pyeon, M. W., Kim, S. W., Kim, J. R., Han, D. Y. (2012). "Coregistration of terrestrial lidar points by adaptive scale-invariant feature transformation with constrained geometry." *Automation in Construction,* 25, 49-58.

Golparvar-Fard, M. , Bohn, J. , Teizer, J. , Savarese, S. , and Peña-Mora, F. (2011). "Evaluation of image-based and laser scanning accuracy for emerging automated performance monitoring techniques*." Automation in Constrction* , 20 (8), 1143–1155.

Fischler, M. and Bolles, R., (1981). "Random sample consensus: A paradigm to model fitting with applications to image analysis and automated cartography." *Comm. ACM*, Vol. 24, No. 6, pp. 381–395.

Huber, D., Akinci, B., Tang, P., Adan, A., Okorn, B. and Xiong, X. (2010). "Using laser scanner for modeling and analysis in architecture, engineering and construction," *In Proceedings of the Conference on Information Sciences and Systems (CISS)*, Princeton, NJ.

Li, Y. and Wang, Y.(2008). "An accurate registration method based on point clouds and redundancy elimination of LIDAR data." *XXI ISPRS Congress: Commission V*, WG 3, Beijing, China, pp605-610.

Liu, S. and Xie, X. (2011). "Research on Algorithm Of Point Cloud MapReduce Registration." *Proceedings of IEEE CCIS2011*, IEEE, Beijing, China, Sep. 15-17, 338-341.

Matas, J., Chum, O., Urban, M. and Pajdla, T.(2004). "Robust wide-baseline stereo from maximally stable extremal regions." *Image and Vision Computing*, 22(10).

Nister, D.(2003)." Preemptive RANSAC for live structure and motion estimation." *In: Proc. ICCV*,vol. 1, pp. 199–206.

Men, H., Gebre, B., Pochiraju, K. (2011). "Color Point Cloud Registration with 4D ICP Algorithm." *International Conference on Robotics and Automation*, IEEE, Shanghai, China, May 9-13, 1511-1516.

Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys. (2008)."A comparative analysis of ransac techniques leading to adaptive real-time random sample consensus." *In Proceedings of the European Conference on Computer Vision (ECCV)*.

Teizer, J., Allread, B. S., Mantripragada, U. (2010). "Automating the blind spot measurement of construction equipment." *Automation in Construction*, 19(4).

.

Modeling the Production Capacity of a Continuous Plant Using Discrete Event Simulation

S. C. Lau[1], M. Lu[2] and C. S. Poon[3]

[1]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[2]Construction Engineering and Management, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta, Canada; PH (780) 492-5110; email: mlu6@ualberta.ca
[3] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

ABSTRACT

Thanks to its simplicity and cost-effectiveness, discrete event simulation is preferred for construction simulation and the issue of how to model the production capacity of a continuous plant by using discrete event simulation is addressed. A plant of continuous nature relies on a material-handling mechanism (such as conveyor or pipeline) to continuously convey material delivered in truckloads to a designated activity location at the site. In contrast with discrete resources commonly encountered and matched in construction (such as a truck, an excavator, and a crane), a buffer is the hallmark of a continuous plant (such as unloading container); and one or multiple feeder resources (trucks) can be simultaneously processed subject to the production capacity of the plant. With a concrete pump example, we discuss the potential pitfall of simplifying a continuous plant as one discrete resource entity. Then, we formalize an approximate method for representing a continuous plant with N discrete resource entities in simulation of a predominantly discrete operations system. We also describe two practical applications: (1) modeling the production capacity of a magnetic separation plant in iron mining operations; (2) modeling the passing capacity of an urban road section in microtunneling and pipe-jacking operations. In conclusion, the proposed method adds to the usefulness and flexibility of a discrete simulation methodology in modeling complicated construction systems.

INTRODUCTION

Simulation modeling builds a logical model on the computer medium as a valid, adequate representation of a complicated problem in reality, aiming at achieving a better understanding of the problem and hence resolving the problem. (Law and Kelton 2000) With respect to the mechanism by which the state of the system changes over time, simulation methodologies can be broadly categorized into discrete event simulation and continuous simulation (Prisker and O'Reilly 1999).

In discrete event simulation, the modeler is concerned with how to describe the logical conditions for triggering the occurrence of events that change the system state only at discrete points in time. The majority of simulation applications in construction engineering fall into the discrete class for its simplicity and effectiveness

(Shi and AbouRizk 1998.) Therefore, discrete event simulation provides the norm viewpoint for representation of a construction operations system into a simulation model. CYCLONE (short for CYClic Operation NEtwork, Halpin 1977), along with its later extensions and add-ons, is the best-known discrete simulation method used in construction engineering research.

As regarded continuous simulation, the state variables of the system change continuously with time and such changes are characterized into a set of differential equations. Simple differential equations can be solved analytically, thus, the values of state variables can be integrated against time based on their initial values. However, solving many complex, continuous models needs resort to numerical analysis techniques (e.g. Runge-Kutta integration) in order to evaluate the state variables at a particular point of time (Law and Kelton 2000.) A continuous simulation application in the construction domain was to model the drawdown of underground water table over time as a result of construction site dewatering operations by a system of pumps (Hajjar et al 1998.)

In simulation of practical construction systems, certain elements−within a predominantly discrete system are continuous in nature −being resources or processes. Modeling such complex systems involves both discrete and continuous simulations, resulting in the hybrid viewpoint of combined simulation (Law and Kelton 2000). The dependent variables may change discretely, continuously, or continuously with sudden jumps, contingent on the occurrence of time events or state events. The key characteristic of the combined simulation paradigm lies in the interactions between system variables with respect to the following aspects: (1) a continuous variable may take a discrete change in value at a time event; (2) an event involving a continuous state variable reaching a threshold value may trigger the occurrence of an event; (3) the functional description of continuous variables may be altered at discrete times (Pritsker and O'Reilly 1999). Commercial discrete-event simulation packages provide the functionality for incorporating continuous elements into a discrete system model. In particular, the SLAM/AweSim system (Pritsker and O'Reilly 1999) has been used for combined simulation in construction research.

In order to estimate the effect of weather on productivity and duration of weather-sensitive activities, a combined discrete-event/continuous simulation was to link the continuous weather parameters with the discrete-event project scheduling model by integrating the use of SLAM simulation platform, MS Project, and NeuralWindows (AbouRizk and Wales 1997). Based on the SLAM simulation platform, a continuous simulation model and a discrete simulation model were independently developed and contrasted for a pipeline project (Shi and AbouRizk 1998). Note the continuous model defines a set of differential equations to represent the continuous progress of consecutive activities on the project, while the discrete operations model depicts the resources' construction cycles in detail. The comparison concluded that the discrete model provides more flexibility while entailing less difficulty than the continuous model.

PROBLEM STATEMENT

In a predominantly discrete construction system, a continuous plant features a material-handling mechanism (such as conveyer or pipeline) that continuously conveys material −delivered by transit resources (such as trucks or truck mixers) − to a designated location in the site. A discrete batch of material is not readily identified and easily observed in the material handling process by a continuous plant. A plant of continuous nature has a limited production capacity in terms of the quantity of material processed in a time unit (hour or day). Often, a continuous plant constitutes the "bottleneck" resource in a site production system, driving the configuration of other supporting resources and controlling the overall productivity performance. For the continuous plant like pump, its production capacity is implicit if discrete modeling is applied.

Let us consider a case of a concrete pump equipped with a feeder container and pipeline, which continuously pumps concrete from the unloading point −where mixer trucks are unloaded− to the concrete-placing point situated on the floor being built. In this case, the concrete pump can be seen as a continuous plant. In contrast, if a tower crane is used to pour concrete, a skip-load of concrete can be readily identified and tracked as a discrete batch of the material being handled. Thus, the material flow in concrete pouring is not continuous, nor is the crane a continuous plant. Other examples of continuous plants include (1) an aggregate production plant with a conveyor system to process truckloads of raw material into aggregates of various sizes in continuous flows; (2) an iron ore processing plant with magnetic separation drums for extracting iron sand from the slurry of iron ore (Lu et al. 2007); and (3) a dual-lane road section for carrying urban traffic flows which include construction trucks delivering precast pipe sections to a microtunneling and pipejacking site in a city (Lau et al. 2010).

COMBINED MODELING APPROACH

Apparently, the "continuous plant" problem can be tackled with a combined modeling approach. On one hand, the plant production rate function is defined for continuous modeling, which is integrated over the simulation time to derive the production output. On the other hand, the production cycle of trucks (arrival, waiting, unloading, returning to batching plant) is modeled by discrete event simulation. A combined simulation executive program seamlessly blends the two simulation paradigms during dynamic execution of the simulation model.

However, the downside of a combined simulation approach resides in the expense of additional time spent in developing a detailed model (AbouRizk and Wales 1997). For instance, beyond developing a diagrammatic model by connecting basic SLAM modeling elements, Shi and AbouRizk (1998) inserted FORTRAN code as the SLAM subroutines written to realize continuous modeling of two repetitive pipeline construction activities. They pointed out modeling resource sharing between activities is less straightforward in a continuous model, and observed "the major modeling functions in a continuous model have to be coded by the user, making continuous simulation more difficult to implement."

Hence, the problem statement for the present research is simple: in a predominantly discrete operations system, is it possible to devise a quick yet valid

method for modeling the production capacity of a continuous plant with discrete resource entities? As such, applying the discrete simulation method (such as CYCLONE) is sufficient and accurate to simulate the complete operations of construction. This would not only add to the usefulness and flexibility of a discrete simulation methodology, but also help reduce the application cost of construction simulation methods in terms of software expenses and learning efforts.

CONTINUOUS PLANT DISCRETIZATION

Herein, a straightforward methodology is proposed for discretizing a continuous plant in a discrete simulation model of construction operations. First, the potential pitfall of modeling a continuous plant by simplifying it as one discrete resource entity is discussed and illustrated with a concrete pump example.

Let us take the modeling of a concrete pump for example: at a building site, truckmixers deliver concrete into the feeder of a stationary pump. Note in contrast with discrete resources commonly encountered and matched in construction (such as a truck, an excavator, and a crane), a buffer is the hallmark of a continuous plant (such as unloading container); one or multiple feeder resources can be simultaneously processed subject to the production capacity of the plant. Concrete is continuously pumped up the pipeline from the unloading point to the placing point on the upper floor. In the development of a discrete simulation model of the entire site operations, one critical issue is how to represent the pump's production capacity in processing truckmixers and placing concrete. An easy way to model the concreting process is to treat the pump as one scarce resource, which is then matched with one mixer truck before engaging in the pumping activity. As shown in Fig. 1, two "Queue" nodes plus one "Combi" activity form a basic CYCLONE model structure, with the "Pump Queue" and the "Mixer Truck Queue" denoting the resource requirements to invoke the pumping activity. One resource entity is initially placed in the "Pump Queue" (as symbolized with an asterisk) to indicate the availability of one pump resource at site.
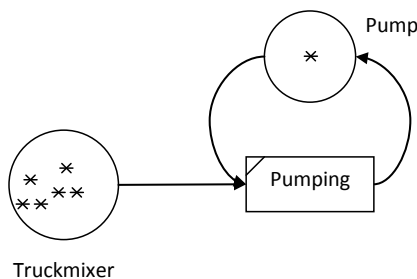

Figure 1. Simple CYCLONE model for concrete pumping process

The "Mixer Truck Queue" is initialized according to prescheduled truck delivery time (shown in Fig. 1), or alternatively, is dynamically linked to the concrete delivery cycle between the batching plant and the building site. Note the quantity of concrete carried in one truckload is known; and the activity time distribution of

"Pumping Combi" models the uncertainty in the processing time required for unloading concrete from one truck at the site. The above CYCLONE model structure is commonly used to represent conveyers in batching plants or pumps in building sites in previous research of concrete-placing operations simulation (e.g. Zayed and Halpin 2000). However, the continuous nature of the concrete pumping operation would likely render the discrete model in Fig. 1 inaccurate, especially when the production capacity of the pump is high. This is explained as follows.

A pump with a rated production capacity of 43 m^3 per hour is used for pouring a bored pile in foundation construction. Truckmixers of 7-m^3 volume capacity each arrive at the site in time, ensuring smooth, continuous concrete pumping. Given the average duration for unloading one 7-m^3 truckmixer is 20 min, we can deduce the simulation result in the first one hour by quick hand simulation based on the simple CYCLONE model as shown in Fig. 1. From the start to the 20th minute, the first truckload is processed by the pump. From the 21st minute to the 40th minute, the second truckload is processed. Then, from the 41st minute to the 60th minute, the third truckload is processed. Over the first hour, the pump resource is 100% utilized and a total of 21 m^3 concrete is pumped. However, an obvious mismatch can be observed between the simulation result and the expected productivity performance with respect to the pump's production. The production rate of the pump is rated 43 m^3 per hour. That means with the pump running at its full capacity in the first hour, the actual quantity of concrete pumped is supposed to be close to six truckloads, i.e. 42 m^3. Yet, executing the CYCLONE model would have only yielded three truckloads (or 21 m^3.) In short, the above simple case has exposed a potential pitfall in the modeling of a continuous plant by applying discrete event simulation. As a matter of fact, the pump has the production capacity of unloading two truckmixers simultaneously; how to model the pump by discrete event simulation is as follows.

To fix the problem, we now designate (1) the production rate of the continuous plant in terms of the quantity of material processed in an hour as P; (2) the event time of "start processing one unit of production (truckload)" as t_1 and the event time of "finish processing one unit of production" as t_2; and (3) the quantity of material delivered by one truckload as q. Here, Eq. (1) is proposed to determine the quantity of "pseudo resource entities" (N), used to initialize the availability of the continuous plant:

$$N = \frac{\int_{t_1}^{t_2} P \cdot dt}{q} = \frac{\overline{P} \times (t_2 - t_1)}{q} \qquad (1)$$

The following explanations are made on the definition of N:
• N is dimensionless and rounded off to the closest integer. Different from the quantity of discrete resources commonly employed in construction (such as equipment, tools or crews), N is not the actual count of physical resource elements but the quantity of "pseudo resource entities" specifically defined to model the production capacity of a continuous plant.
• The numerator is the integration of the production rate (P) of the continuous plant over the time period [t_1 ,t_2]. The difference of (t_2-t_1) can be taken as the average time duration for the continuous plant to process one production unit while $\overline{P}$ is taken as the average production rate of the continuous plant over that period. On the

denominator, q is the quantity of material contained in one production unit (truckload). As a production unit represents the amount of material contained in one truckload, given a plant that handles the continuous flow of production units, the quotient N can be visualized as the maximum number of channels within the plant that allow the production units to flow in parallel.

•   N is approximated as a constant based on the average or most likely values of , t_2-t_1, and q. Variability in those parameters due to random variations or uncertain site factors can be conditioned into the statistical distribution of activity time, which is used for Monte Carlo sampling during simulation.

For the above concrete pump example, the production unit is one truckload of concrete. $\overline{P}$, t_1, t_2, and q are 43 m^3 per hour, 8:00 am, 8:20 am, and 7 m^3, respectively. Thus, N is decided to be 2:

$$N = \frac{\overline{P} \times (t_2 - t_1)}{q} = \frac{43m^3 / hr \times (8:20 - 8:00)\min/60}{7m^3} \approx 2$$

To update the CYCLONE model (Fig. 1), two resource entities are placed initially at the queue node associated with the pump's availability. Then, repeating hand simulation on the CYCLONE model would result in six truckloads (or about 42 m^3) of concrete being processed and 100% utilization for the two pseudo resource entities in the first hour (denoting full utilization on the pump). In the event of activity interruption due to concrete supply problems or other site factors, the pump would not operate at its full capacity. For instance, tardy concrete deliveries would cause interruptions to the pumping process, thus reducing the utilization rate of the two pseudo resource entities to 70%. This would bring down the pump production in the first hour from 42 m^3 (six truckloads) to 28 m^3 (four truckloads).
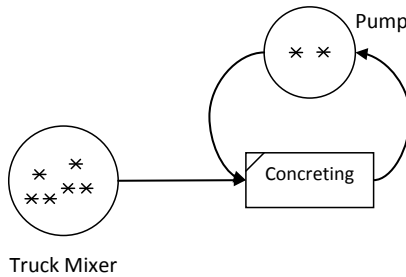
Figure 2. Updated CYCLONE model with two resource entities being placed initially at the queue node associated with the stationary pump

Next, we present practical applications of defining "pseudo resource entities" to model (1) the production capacity of a continuous plant in mining operations, based on our experience of utilizing discrete event simulation to facilitate the design of production facilities for an iron ore mine situated in Indonesia; and (2) a dual-lane urban road section for carrying traffic flows which include construction trucks delivering precast pipe sections to a microtunneling and pipejacking site in Hong Kong.

PRACTICAL APPLICATIONS

Two practical application cases are described: 1) modeling the production capacity of a magnetic separation plant in iron mining operations; (2) modeling the passing capacity of an urban road section in microtunneling and pipejacking operations.

The production capacity of a magnetic separation plant in iron mining operations is modelled. The plant used in the actual site operations is shown in Fig. 3. The raw sand flurry flows through a series of magnetic drums continuously. As output from the processing plant, the iron sand is separated from waste sand and stone.

Figure 3. Raw material slurry undergoing magnetic separation

Figure 4. TTA setup at the middle of a road section

The production rate (P) of the magnetic iron sand separation plant is designed as 140 ton per hour; the iron sand contained in one truckload (q) is 4.12 ton (i.e. one truckload carries 19 ton of raw material, 15% of which is iron sand while the remaining 85% is waste); the time duration required for the processing plant to handle one truckload (t_2-t_1) was observed to be 9 minutes on average. N is determined as:

$$N = \frac{P \times (t_2 - t_1)}{q} = \frac{140 ton / hr \times 9 \min/ 60}{4.12 ton} \approx 5 \qquad (3)$$

Hence, the magnetic separator plant can be modeled with five "pseudo resource entities" in a discrete simulation system. Interested readers may refer to Lu et al. (2007) for detailed complete simulation modeling and application.

The passing capacity of an urban road section relevant to microtunneling and pipe-jacking operations can also be modelled using the proposed method. A typical urban road section next to the jacking shaft with Temporary Traffic Arrangement (TTA) is taken as a continuous plant, shown in Fig. 4. L_i is the length of the lane (100 m) near the jacking shaft; V_p is the vehicle speed (30 km/hr). The traffic passing capacity of the road section is estimated to be 1200 vehicles/hr. In this case, the carrying capacity of the road section can be modeled with N resources in Eq. (4), which is a variation defined based on Eq. (1); Note q is simply 1 vehicle

$$N = \frac{P \times (t_2 - t_1)}{q} = \frac{P \times L_i / V_p * 60 * 60 / 1000}{q} \qquad (4)$$

$$= \frac{1200 veh / 3600 \sec \times (100 m/30 km/hr \times 60 \sec/\min \times 60 \min/hr/1000 m/km)}{1 veh} \approx 4$$

Therefore, the traffic carrying capacity for a particular road section of 100m can be determined N = 4. Interested readers may refer to Lau et al. (2010) for detailed traffic and construction modeling.

CONCLUSION

This research has addressed the issue of how to model the production capacity of a continuous plant by using discrete event simulation. A plant of continuous nature relies on a material-handling mechanism (such as conveyer or pipeline) to continuously convey material delivered in truckloads to a designated activity location at the site. We have formalized an approximate method for representing a continuous plant with N discrete resource entities in simulation of a predominantly discrete operations system. N is the quantity of "psedo resources" with no physical meaning, only representing the continuous plant would process N feeder resources simultaneously if the plant were modeled by DES, as such the passing capacity of the plant matches the simulation result. We also describe two practical applications to demonstrate the usefulness and flexibility of a discrete simulation methodology in modeling complicated construction systems.

REFERENCES

AbouRizk, S. and Wales, R. 1997. "Combined Discrete-Event/Continuous Simulation for Project Planning." *Journal of Construction Engineering and Management*, ASCE, Volume 123, No. 1, pp. 11-20.

Halpin, D. W. 1977. "CYCLONE – method for modeling job site processes." *J. Constr. Div*., ASCE, Vol. 103, No.3, 489-499.

Hajjar, D., AbouRizk, S., and Xu J. 1998. "Construction site dewatering analysis using a special purpose simulation-based framework." *Canadian Journal of Civil Engineering*, 25, 819-828.

Lau, S.C., Lu, M., Ariaratnam, S.T., et al. 2010. "Simulation-Based Approach to Planning Temporary Traffic Arrangement for Microtunneling Operations in Urban Areas." *10th International Conference on Construction Applications of Virtual Reality 2010*, 4-5 November 2010, Sendai, Miyagi, Japan,.

Law, A.M. and Kelton, W.D. 2000. *Simulation modeling and analysis*, 3rd ed. McGraw-Hill, N.Y.

Lu, M., Lau S.C., and Chan E.K.Y. 2007. "Combined simulation modeling using simplified discrete event simulation approach −a mining case study." *The Proceedings of the 2007 Summer Computer Simulation Conference*, pp. 421-428, the Society of Modeling and Simulation, San Diego, CA, July 2007.

Pritsker, A. A. B. and O'Reilly, J. J. 1999. *Simulation with Visual SLAM and AweSim*, 2nd ed. Systems Pub. Corp., West Lafayette, Ind.

Shi, J. and AbouRizk, S. 1998. *"Continuous and combined event-process models for simulating pipeline construction"*, Construction Management and Economics, Volume 16, 489-498.

Zayed T.M. and Halpin D.W. 2000. "Simulation as a tool for resource management." P*roceedings of Winter Simulation Conference 2000*. pp.1897-1906, Vol.2, Winter Simulation Conference Foundation, Orlando, FL, USA.

Exploring Shelter Logistics and Infrastructure Attributes during an Extreme Event using Agent-Based Modeling

Vanessa Valentin[1], Nader Naderpajouh[2] and Freddy Solis[3]

[1] Assistant Professor, Department of Civil Engineering, University of New Mexico, MSC01 1070, Albuquerque, NM 87131, USA, Phone: 505-.277-.0811, vv@unm.edu

[2] Ph.D. Candidate, School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 4790947907-2051, USA, Phone: 765 -496-2046, nnp@purdue.edu

[3] Ph.D. Student, School of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 4790947907-2051, USA, Phone: 765 496-494-/765-464-93233974, fsolisno@purdue.edu

ABSTRACT

Extreme events have always imposed radical constraints on planners and logistic managers. Fast paced dynamics of these events necessitate more deliberate review of response plans. One aspect of extreme events for decision makers and policy developers is the potential mobilization of the victims in the affected community and logistics of shelter provision. Within this focus area, research that integrates infrastructure characteristics is limited. This study presents an Agent-Based Model (ABM) to evaluate four different scenarios of community mobilization and shelter provision logistics during an extreme event: (1) victims go to a random shelter, (2) victims are directed to their closest shelter, (3) victims are mobilized based on a centralized shelter designation, using "triages" and, (4) the shelter designation and the mobilization of victims is performed considering pre-established zones. The model integrates the exchange of information between agents (community and authorities). Variables such as capacity of transportation infrastructure, number and location of shelters, number of victims, among others, are also incorporated into the model. Preliminary results indicate the potential for application of ABM to improve the understanding of interactions among agents, their associated dynamics, and their impacts on different performance measures of the system (i.e., number of victims accommodated into shelters, and accommodation time) during an extreme event. Scenarios that incorporate triages did not always provide the best shelter provision results.

Introduction and background

Extreme events are low probability high consequence (LP-HC) events (Kunreuther et al. 2004). The consequences of any non-catastrophic event comprise destruction of facilities used by multiple systems, such as infrastructures, or interruption of routine activities of in affected systems. In case of extreme event, however, the magnitude and impact of destructions and interruptions are substantially

larger. The consequences also prolong longer than in non-catastrophic events, augmenting the importance of effective and efficient large-scale responses (USDOT, USDOHS, 2006).

Extreme events have a broad array of causes, which span from natural to man-made (Kunreuther et al. 2004). Because of the diversity of these events, as evidenced throughout many historical extreme events and their causes, different scenarios require different response policies. Examples of extreme events with natural origin include (but are not limited to): hurricanes, tornados, floods, earthquakes, wildfires, and volcano eruptions. Examples of man-made extreme events include (but are not limited to) terrorism, dam failures, and nuclear power plant emergencies (FEMA 2009). Each category of extreme event has its own scenarios with specific characteristics, which necessitates application of specific policy and logistic response plans. Though existing plans and logistics in affected regions are typically adequate for non-catastrophic events, such plans might not address constraints imposed by extreme events (USDOT, USDOHS, 2006).

Problem Statement

Extreme events are multifaceted challenges that encompass diverse interrelated issues. Two aspects of extreme events that are common to involve mobilization of the public and shelter provision (in this paper, mobilization refers to the random change in the location of the people, since transportation is by itself a different aspect of the problem). For example, the Federal Emergency Management Agency (FEMA) conducted Hurricane Evacuation Studies (HES) comprising five major analyses: hazard analysis, vulnerability analysis, behavioral analysis, shelter analysis, and transportation analysis (Massey 2008). Shelter analysis is defined as the detailed inventory and assessment of public shelter facilities and their accessibility for cases of extreme events (Massey 2008). Since shelter provision is an elemental segment in both short-term and long-term response and recovery process (Bolin and Stanford 2007), the research question of the study can be defined as follows:

What are the consequences of different shelter provision logistics, subsequent response policies, and interactions of agents during an extreme event considering infrastructure characteristics?

The study will also investigate relevant performance metrics that should be addressed in evaluation of the shelter provision missions.

Methodology

Agent-Based Model

An Agent-Based Model (ABM) consists of a collection of autonomous decision-making entities called agents, which make decisions according to a set of rules (Bonabeau 2002). According to Bonabeau (2002), the use of ABM is well suited for multiple application areas including flow analysis. In the case of disaster management, Zhang et al. (2009) applied ABM for hurricane evacuation focusing on the transportation aspects – which is relevant to extreme events. This study in particular focuses on the distribution of victims to shelters considering three agents:

(1) triages, (2) shelters and, (3) victims. An ABM was developed to simulate, under different cases, the interactions between these agents during an extreme event and to understand the dynamics and impacts of these interactions on the performance measures (i.e. the number of victims accommodated in shelters, accommodation time and cost) of the process of shelter provision.

Cases

Four different cases were studied in order to answer the proposed research question. The first two cases developed according to the existing policies for shelter provision. No triages are modeled in these two cases.The rest of the cases investigate the use of triages in shelter provision in order to explore the impact of this organization in the shelter provision process. The modeling process started with a minimal set of parameters and after developing a richer understanding of the problem the complexity of the problem was increased by adding more parameters to achieve a more realistic model of shelter provision mission.

- Case #1: Victims go to closest shelter. If the closest shelter is full they will go to the second closest shelter. In this case, victims have full information on the location of the closest shelter. This information could be obtained by scanning the media or informative booklets prior or during the extreme event. Shelters will have information about their maximum and remaining capacity, and also about the location of the closest shelter. In this case, it is assumed that there is no exchange of information between the shelters, but there is exchange of information between victims and shelters (i.e., remaining capacity of shelter) in case they come to the shelter. However, the shelters provide information to victims regarding the location of other shelters. It can be said that the system has only one network (network of shelters), while there is no exchange of information during the mission among them. Depending on the location of the shelters and density of region, some emergent behaviors such as traffic congestion may be expected.

- Case #2: Victims go to a random shelter. Then, if this shelter is full they will go to the closest shelter. The shelter will provide this information to the victim. In this case, it is assumed that at the initial stages of the mobilization victims do not have information regarding the location of the closest shelter. This case aims to prove the efficiency of information sharing before the disaster. The shelter network, as a different system, will operate same as the case #1. The difference lays in the knowledge of victims, since they do not know their closest shelter. The case exhibits further independent operation and management of shelters as well as the role of information sharing before disaster in the efficiency of the mission.

- Case #3: Victims will go to the closest triage. The triage will send people to the closest shelter until the shelter is full. In this case, it is assumed that there is an exchange of information between (1) shelters and triages (i.e., remaining capacity), (2) victims and triages (i.e., the location of the closest shelter with

the remaining capacity), and between triages themselves (i.e., remaining capacity of the shelters). Shelters do not exchange information between them. The congestion variable was added into the model to reflect more realistic behavior of the shelter provision mission and its subsequences. Emergence behavior of gridlock, in case the ratio of triage to victims is low, may be observed.

- Case #4: Victims are mobilized to triage according to the zone they are located. The mobilization of people occurs in one zone at a time. The information exchange between agents is the same as in Case #3.

Model

NetLogo (version 4.0.4) was used to develop the ABM that simulates the interactions between the agents involved in the process of shelter provision during an extreme event.NetLogo is an ABM Software developed by Northwestern University characterized by its simplicity and accessibility among various software tools for creating ABMs. NetLogo provides an extensive library of example models, including social network models, which is related to our shelter provision system.

The modeling space of the ABM consisted of a grid of 100 x 100 patches. This space represents the city in which the extreme event is occurring or is about to occur. The model was developed so that it could be easily adapted to the policies of different types of extreme events.

The model interface (shown in Figure 1) cases consists of multiple slide bars for each one of the cases, used to input the different parameters to run the simulation. These parameters include: number of victims, number of shelters, capacity of shelters, number of triages, resources at triages, resources at shelters, and congestion. The resources represent the number of people who can be served by the triage/shelter at the same time. It investigates the performance of different cases considering,, queue of victims formed outside the triage. The developed model did not include the physical simulation of available infrastructure within the target city. However, the congestion parameter was included in order to reflect the infrastructure characteristics of the city to a certain extent. This parameter establishes a maximum number of people that can be in a place at a time and it is similar to the term capacity in traffic analysis. Congestion in the traffic context occurs when traffic flow approaches roadway capacity (Fricker and Whitford 2004).

In addition, the model interface includes a visual representation of the interactions of the agents and a plot that is updated with each time step. This plot provides valuable information to analyze dynamics between agents. For instance, it provides the number of victims accommodated into shelters as well as people waiting in triage at a specific time.

The modeling process started from a simple Case with minimal complexity (e.g. model without congestion). As complexity was gradually added to the model, key behaviors were observed in different cases to check compatibility of the computer model with the base paper model (e.g. victim mobilization from triage to candidate shelters). The review included examination of the independent components of the system. Operational validity is difficult to achieve at this stage of this study.
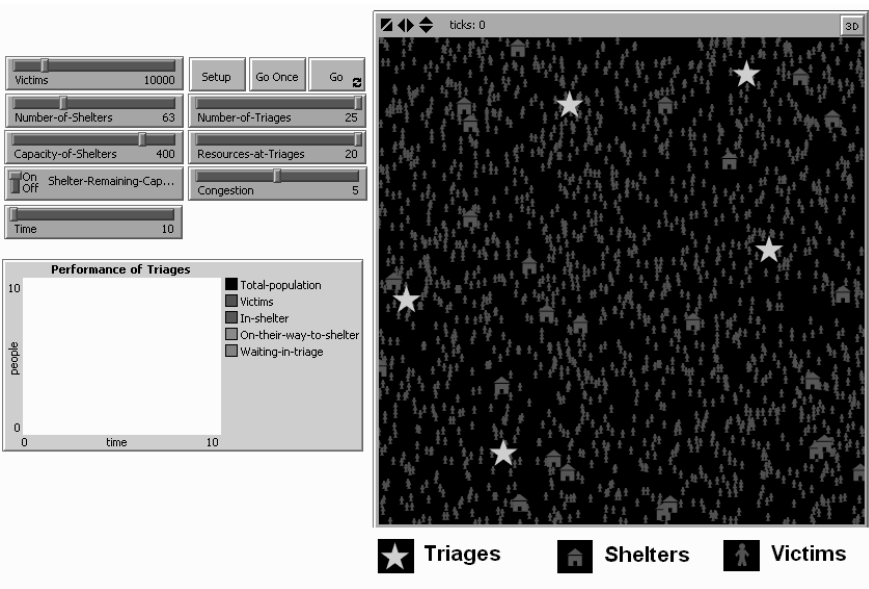
Figure 1. Initial program interface for shelter provision models

Analysis of Cases

Cases #1 through 4 were compared in terms of the number of people accommodated in shelters during the response period. The response period is defined as the first 50 hours since the start of the mobilization towards shelters. Each case took approximately from 15 to 20 minutes to run. In order to evaluate the performance of the different cases, the values of the parameters were considered similar. The assumptions made for each parameter are the following:

- Number of victims: 30,000
- Number of shelters: 100
- Capacity of shelters: 300
- Number of triages: 15
- Number of resources at triages: 15
- Number of resources at shelters: 5
- Congestion: Medium (5 in a scale from 5 to 10)

In case #1, victims were being accommodated into shelters very rapidly at the beginning of the simulation. This happens because at the beginning of the simulation the shelters are completely empty, and once victims reach the closest shelter they are accommodated. However, once the shelters are getting full, people start moving around looking for a shelter with the remaining capacity (since they just have information about the location of the next closest shelter and not about the capacity of

the shelters). At this stage accommodation continues at a lower rate. The peak of people waiting at the triage occurs after 20 hours.

In case #2 victims are being accommodated into shelters at a constant rate. Since the victims do not have information on the location of the closest shelter, they initially go to a random shelter that is not necessarily the closest one. People waiting to be accommodated at the shelters are also increasing in a linear way.

Case #3 considers the use of triages with the incorporation of the "congestion" parameter in order to account for infrastructure characteristics (in this case transportation). While running the model, most of the people never get to reach the triage, since the capacity of the infrastructure was met. This is comparable to what would happen when there are gridlocks or traffic jams in the transportation network. One possible solution to this problem could be to conduct people to shelters according to the area they are located. This particular case is case #4.

Based on the number of mobilized victims, case #4 provided the best option for victim mobilization. Victims did not encounter high congestion as in the case #3. In addition, the wait time at the shelters was less than the previous cases.

Since the initial location of the victims and location of the triages and shelters are randomly generated in the model, each case was run several times and then the data was averaged. Figure 2 summarizes and compares the results for each case. Considering model parameters, the largest number of victims that were accommodated at 50 hours occurred in case #4 (when zones were designated). These results emphasize the vitality of proactive advertisement regarding the location of the closest shelter to citizens before the disaster event.

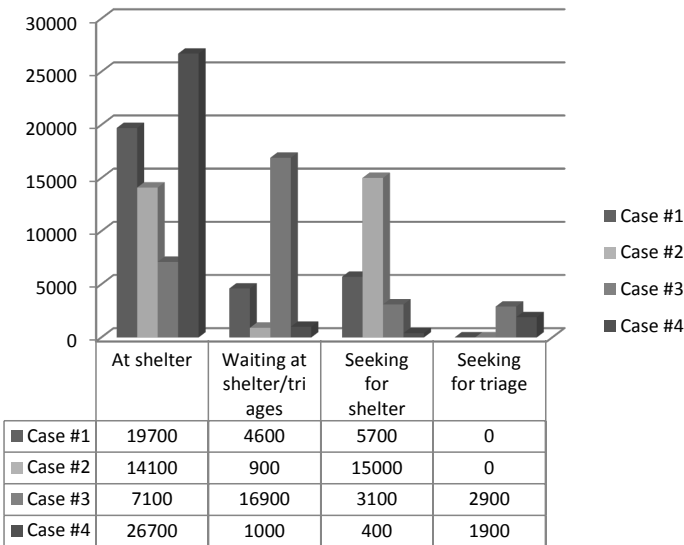|	At shelter	Waiting at shelter/triages	Seeking for shelter	Seeking for triage
Case #1	19700	4600	5700	0
Case #2	14100	900	15000	0
Case #3	7100	16900	3100	2900
Case #4	26700	1000	400	1900

Figure 2. Summary of results for each case

Conclusions, Limitations and Future Work

Providing shelter to victims during extreme events involves interaction of multiple independent agents. In this study, an ABM was developed to simulate and understand, under different cases, the interactions between these agents during an extreme event and to understand the dynamics and impacts of these interactions and logistics. According to the models that were developed and the assumed parameters, the cases in which triages were incorporated did not always provide the best shelter provision results, which may be attributed to the random location of the triage. In addition, infrastructure restrictions (i.e., congestion) have an enormous impact on the performance of the shelter policies in case of an extreme event. According to the results, it is vital that citizens are provided with information about the shelter provision policies.

One limitation of this study is that it does not evaluate different shelter provision logistics for a physically defined space which could be addressed in future research for verifying the reliability of the results. The study, however, shows that ABMs can be used to improve the understanding of the interactions of agents during an extreme event and to understand the dynamics and impacts of these interactions on different performance measures of the system (i.e., number of victims accommodated into shelters). In addition, through simulating the use of triages the study evaluates the efficiency when using a more centralized approach to shelter logistics, which is limited in existing literature. Future work should consider historical data to validate the model for a specific region. In terms of shelter provision logistics, the following cases can be considered:

-   Given a shelter distribution find the most optimal number of triages and their locations.
-   Using the same performance measures, under which conditions (i.e., number of victims, number of shelters, location of shelters, etc.) the use of triages is warranted and how it compares with the cases in which no triages are use.
-   Instead of victims being considered as individuals, each unit could be considered as a family.
-   Adding more constraints to model infrastructure and their limited capacities.
-   Integration of more behavioral factors into the role of victims.

References

Bolin, R., & Stanford, L. (2007). "Shelter, Housing and Recovery: A Comparison of US Disasters." Disasters, 15(1), 24-34.

Bonabeau, E. (2002). "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems." *PNAS*, Vol. 99, Suppl. 3, 7280-7287.

Fricker, J. D., Whitford, R.K. (2004). "Fundamentals of Transportation Engineering: A Multimodal Systems Approach." Pearson Prentice Hall. School of Civil Engineering. Purdue University.

Kunreuther, H., Meyer, R., Van den Bulte, C., (2004) "Risk Analysis for Extreme Events: Economic Incentives for Reducing Future Loses." National Institute of Standards and Technology .

Massey, W., (2009) "Modernized Hurricane Evacuation Studies-Taking The Traditional Hurricane Evacuation Study To A New Level." Solutions to Natural Disasters Congress, ASCE.

NetLogo 5.0.3 User Manual. <http://ccl.northwestern.edu/netlogo/docs/> Accessed: March 2013.

U.S. Department of Transportation (USDOT) in cooperation with U.S. Department of Homeland Security (USDHS). (2006). "Catastrophic Hurricane Evacuation Planning Evaluation." Report to Congress.

Zhang, B., Chan, W. K., & Ukkusuri, S. V. (2009). "Agent-Based Modeling for Household Level Hurricane Evacuation." *Simulation Conference (WSC), Proceedings of the 2009 Winter* (pp. 2778-2784). IEEE.

Real-time 3D Visualization of Heavy Construction Equipment Operations Using LADAR

Mengmeng Gai[1]; Yong K. Cho A.M.ASCE[2]; Qinghua Xu[3];

[1]GRA, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska - Lincoln, 1110 S. 67th St. Peter Kiewit Institute (PKI) 118A, Omaha, NE 68182, USA; PH +1-402-206-7617; FAX +1-402-554-3850; Email: mengmeng.gai@gmail.com;
[2]Associate Professor, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska-Lincoln, Email: ycho2@unl.edu;
[3]GRA, Charles Durham School of Architectural Engineering and Construction, College of Engineering, University of Nebraska - Lincoln, Email: qinghua.xu@huskers.unl.edu

ABSTRACT

This paper introduces a real-time visualization method using a hybrid LADAR system, to simultaneously assist multiple heavy equipment operators to perceive 3D working environments at dynamic construction sites. A custom-designed LADAR (Light Detection and Ranging) system was used to obtain 3D point cloud data, which can be transferred into data streams and uploaded to a remote operator's computer screen immediately. The data streams can be accessed by different equipment operators through local wireless or global networks such as Wi-Fi and 4G, and be further presented in dynamic 3D views. Strategies to rapidly updating 3D point cloud scenes are discussed in this paper as well. The methodology presented in this paper has been implemented and tested at an equipment operation site and the proposed 3D visualization technique and data transmission network system were validated effectively. It is highly expected that the assistance of 3D visualization for equipment workspace will significantly reduce blind spots and obstructions, and improves operators' working efficiency.

INTRODUCTION

Safely operating heavy construction equipment such as cranes, excavators, and concrete pump trucks has been considered a very important issue in construction fields. Equipment related fatalities often caused by vision-block problems such as "blind spot" and "blind lift" due to the massive and cluttered working environment. It would be helpful for the operators to access the accurate 3D position of the target objects and surroundings in real-time or near real-time. The development of 3D spatial modeling technology such as laser scanning and machine vision facilitates

equipment control, thus significantly improve construction safety (Teizer et al. 2010). To avoid unpredicted physical contact between a target object and moving equipment, more realistic and accurate data is needed to present spatial perception of the workspace to remote operators (Cho et al. 2002 & 2012). To rapidly process tens of thousands of range data in real-time, which is a critical issue in spatial modeling process, is still an unsolved problem requiring further investigation (Gong and Caldas 2008). Unstructured work areas like construction sites containing highly unpredictable activities and changing rapidly are difficult to visualize graphically. To obtain real-time or near real-time information of surrounding environment further complicates graphical modeling and image updating. Material handling and lifting not only requires rapid 3D visualization, but also accurate position data. One commonly used method to obtain 3D position of the objects is based on 3D scanning technology (Tang, et al. 2010; Huber, et al. 2010), which however has some limitations, such as low speed and low object recognition rates (Kim et al. 2011).

While rapid workspace modeling is essential to effectively control construction equipment (Lee et al. 2009), few approaches have been accepted by the construction industry due to the difficulty to address the visualization challenges under current sensor technologies. In this paper, a LADAR-based rapid workspace modeling method is introduced to assist the equipment operators with real-time 3D scene data transferred and presented on screen inside vehicle cabin. The integrated system was mounted on a moving platform to move around the equipment to provide the equipment operator with better spatial information of blind spots. A field test was conducted at a crane loading site to validate that the proposed 3D visualization method can significantly reduce blind spots and obstructions, and improve operators' working efficiency.

OBJECTIVE

The main objective of this research is to design, develop, and validate a 3D visualization method to measure and process spatial information rapidly in cluttered construction job sites for the effective construction equipment operations. To realizes that, a hybrid sensor system and modeling algorithms were designed to collect field data of equipment operation. The visualization method utilizes instant sensory data from the hybrid LADAR system to implement automatic workspace visualization in a 3D working environment. A real-time data transmission via wireless technology was implemented and tested to transmit and display point cloud and image data to the remotely located computer screen at an equipment operation cabin.

DEVELOPED DATA ACQUISITION HARDWARE SYSTEM

From the previous research effort, an innovative robotic hybrid LADAR system was developed, consisting of two 2D line laser scanners (40 and 80 meters working ranges at 100Hz scan speed, up to 2.5 sec / 360° scan, 190° for vertical line), a digital video camera (2592 x 1944 at 17Mbps/VBR), and a digital camera (1280 x 720 pixels at 30 fps), as is shown in Figure 1. The resolution is 0.125 degree in

vertical direction and 0.18 degree in horizontal direction, with a rotation speed of 1.08 degree per second. A graphical user interface (GUI) was built using Visual C ++. The GUI controls the LADAR scanner and cameras, and visualizes 3D point clouds.
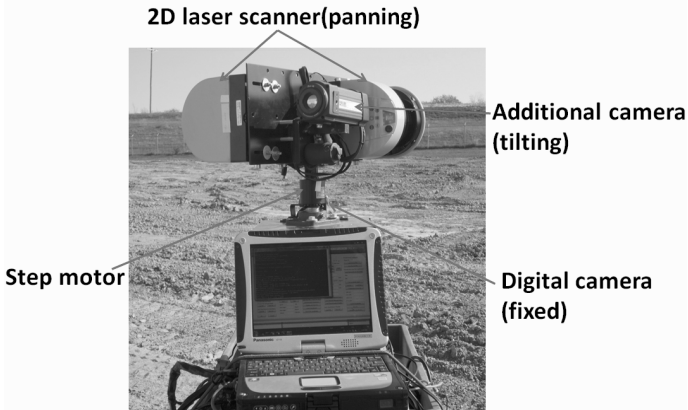

Figure 1. Prototype hybrid LADAR system

Our previous studies have proved that the customized 3D LADAR system can provide more flexibility in hardware control and software programming than a commercial LADAR scanner (Gai et al. 2012; Im et al. 2012; Cho and Martinez 2009). Based on the mounting configuration, multiple degree-of-freedom (DOF) kinematics was solved to obtain x-y-z coordinates from the LADAR, and real-time digital image data were generated from the web camera simultaneously. The transformation matrices for the LADAR and the web camera share same base-coordinate, located in the axle center of the step motor. This kinematics frame allows more sensors such as a digital video camera or a digital camera to be added. The combined two line laser scanners in the hybrid LADAR system allow twice faster scanning rate and twice higher scanning resolution than a single laser scanner. In addition, when applied two laser scanners, the angular resolutions become half of the current minimum degrees by filling vertical and horizontal gaps of point grids.

DATA COMMUNICATION TECHNOLOGIES

Wireless communication technology is preferred for data transmission due to its fast transmission rate and relatively low installment and maintenance costs. Wireless communication is the transmission of data through radio waves instead of any means of physical connections (Amy and Samuel.2013). The wireless system consists of a signal source, a transmitter, a receiver, and two antennas each for transmitter and receiver. Radio transmission is the earliest wireless technology designed by M.G. Marconi in 1901 and is still popular in our nowadays due to its good applicability in voice transmission. However, its relatively low bandwidth for

data limits its capability to perform other forms of data like images and videos, etc. Microwave transmission is another wireless technology which uses high-frequency, short-wavelength radio waves to transport data. Infrared transmission utilizes infrared radiation (an electromagnetic energy) as the medium for data transportation. Because both microwave and IR transmission operate in line of sight mode, they do not perform well through buildings and walls. Therefore, in this study, Wi-Fi and cellular (4G or 4G Long Term Evolution) transmission methods are considered for 3D visualization data transmission.

Wi-Fi communication was recognized by IEEE in 1998 as the 802.11 standard that is a framework around local area network (LAN) (Amy and Samuel.2013). Users can receive data at their workstation such as personal computer or laptop within the local area network. There are several advantages of Wi-Fi technology. Firstly, Wi-Fi can host wireless LANs where cables cannot work, such as the outside area and some particular buildings. Many devices can share the same Wi-Fi network in the covered area, and the price of the equipment needed by Wi-Fi network continues drop. Wi-Fi employs a secured strategy named Wi-Fi Protected Access encryption (WPA2) and power saving mechanisms to extend battery life. The fourth generation of wireless (4G) communication is the latest communication technology that provides secure mobile access to smartphones. Remote users can access the data from smartphones or any mobile device at their convenience.

REAL TIME DATA TRANSMISSION AND VISUALIZATION

The raw 3D spatial measurement data will be accurately and successfully transferred to the desired remote computing system by the selected communication technologies. Among various available techniques for wireless data communication, Wi-Fi and 4G are utilized in this study to establish the connection between a remote computer and the hybrid LADAR system. In the local area, the decision makers (e.g., equipment operator) can access the 3D data through the Wi-Fi based wireless local area network (WLAN). In a remotely located area, the data transfer can be implemented through 4G based wireless wide area network (WWAN). In our study, the 3D point clouds and image data obtained from the hybrid LADAR system are transferred into data streams and uploaded to computer screen of a remote equipment operator. Figure 2 illustrates the system configuration, including the hybrid LADAR system, data server, router and terminals.
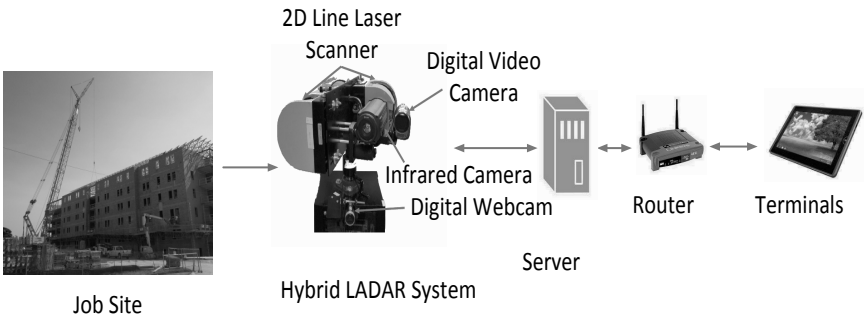
Figure 2. Data communication system configuration

Mounted on a mobile platform (e.g., wheeled cart), the hybrid LADAR system gathers point clouds of a working space. A separate data server connected to the LADAR system automatically stores the scanned data set and shares the data through a wireless router using Wi-Fi technology. Different mobile terminals can access the data server across the wireless router and present 3D model of the working space. Multiple remote equipment operators or project managers can also access the 3D model data from any mobile device at their convenience via 4G technology.

To validate the proposed methods, a crane loading site was tested. In this field test, the data transferring rate between the LADAR system and terminals was about 25 Mbps, the rotation speed of the LADAR system was 7.2 Degree/sec, the resolution in horizontal direction for two laser scanner was 0.42 degree, the resolution in vertical direction was 0.25 degree, the scan angle in horizontal direction was 180 degree, and the scan angle in vertical direction was 190 degree.

Figure 3 (a) shows a workspace including a telescoping mobile crane, a material and an operator, (b) shows 3D point cloud scene obtained by the LADAR system, and (c) shows 3D point cloud model mapped with corresponding texture data capture by a digital camera. Equipment operators can access the data produced by the LADAR system via mobile terminals and investigate the real-time situation of surroundings. The developed system continuously provided time-elapsed 3D scenes of the whole working environment to assist the equipment operator to perceive 3D working spatial information from different viewpoints.
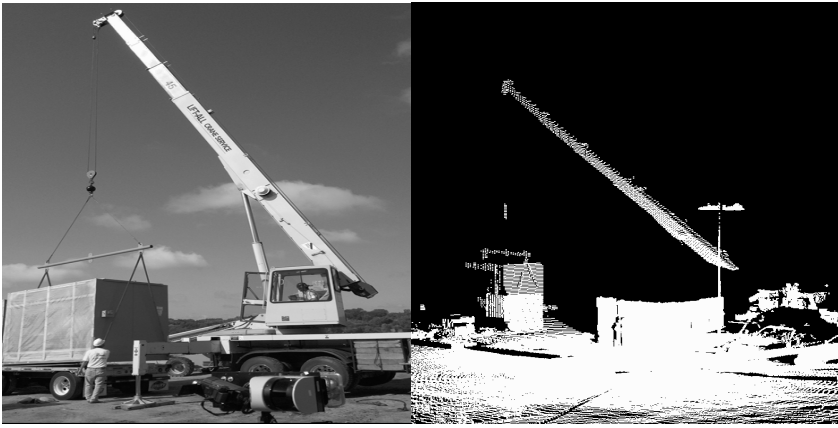
(a) Crane loading site   **(b) Laser scanned point cloud**
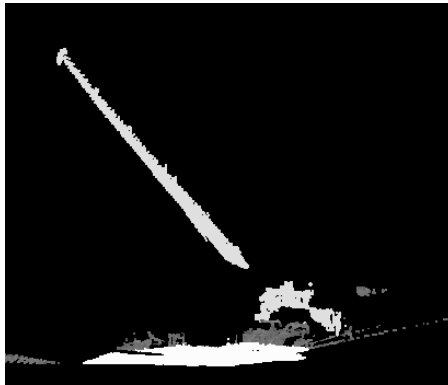
(c) Texture-mapped point cloud

Figure 3. Field experiment results

CONCLUSION

This paper introduced the real time data transmission and visualization methods which allow the equipment operators to perceive 3D working environments at dynamic construction sites. The equipment operator was able to timely access virtual 3D model of their equipment and surroundings via wireless networks. Further, this research demonstrated the merits of visualization technique that multiple equipment operators can share the created 3D site models and remotely monitor equipment operation in real time. The potential use of 3D visualization of working environment will better assist equipment control and improve operation efficiency and safety. Future work will focus on more advanced attempts to combine CAD models of

target objects and 3D reconstructed image model with point cloud scenes to achieve dynamic control of job site.

REFERENCES

Amy, D., and Samuel D. (2013). *Introduction to Low Voltage Systems*, 2nd ed. Clifton Park, NY: Delmar, Cengage Learning, 432-447.

Cho, Y., Haas, C. T., Liapi, K., and Sreenivasan, S. (2002). "A framework for rapid local area modeling for construction automation." *Automation in Construction*, 11(6), 629–641.

Cho, Y., and Martinez, D. (2009). "Light-weight 3D LADAR system for construction robotic operations." *26th International Symposium on Automation and Robotics in Construction (ISARC)*, Austin, Texas.

Cho, Y., Wang, C., Tang, P., and Haas, C.(2012). "Target-focused Local Workspace Modeling for Construction Automation Applications." ASCE Journal of Computing in Civil Engineering, 26(5). pp. 661-670.

Hill, F. S., and Kelley, S. M. (2006). "Rendering Faces for Visual Realism." Computer graphics: using OpenGL, Pearson Prentice Hall, Upper Saddle River, NJ.

Huber, D., Akinci, B., Tang, P., Adan, A., et al. (2010). "Using laser scanner for modeling and analysis in architecture, engineering and construction." *Proceedings of the Conference on Information Sciences and Systems (CISS)*, Princeton, NJ.

Im H., Gai, M., Wang, C., and Cho, Y.(2012). "Hybrid approach to Visualize Building Energy Information Model in Google Earth." *Construction Research Congress 2012*, ASCE, West Lafayette, IN.

Gong, J. and Caldas, C. H. (2008). "Data Processing for Real-Time Construction Site Spatial Modeling." *International Journal of Automation in Construction*, 17 (5), 526-535.

Gai, M., Cho,Y. and Wang, C.(2012). "Projection-Recognition-Projection (PRP) method for object recognition from a 3D point cloud." *Proceedings of the 2012 ASCE International Workshop on Computing in Civil Engineering*, Clear Water, FL.

Kim, C., Lee, J., Cho, M., and Kim, C. (2011). "Fully automated registration of 3D CAD model with point cloud from construction site." *28th International Symposium on Automation and Robotics in Construction*. Seoul, Korea. pp: 917-922.

Lee, G., Kim, H., Lee, C., Ham, S., Yun, S., Cho, H., Kim, b., Kim, g., and Kim, K.(2009). "A laser-technology-based lifting-path tracking system for a robotic tower crane." *Automation in Construction*. 18(7) , pp. 865-874.

Occupational Safety and Health Administration (2009). "Inspections within Industry." <http://www.osha.gov/SLTC/etools/construction/struckby/mainpage.html   > (May 11, 2011).

Tang, P., Huber, D. Akinci, B. Lipman, R. et al. (2010) "Automatic Reconstruction of As-build Building Information Models From Laser-Scanned Point Clouds: A Review of Related Techniques." *Automation in Construction*, vol. 19, pp. 829-843.

Teizer, J., Allread, B. S., Fullerton, C. E., and Hinze, J. (2010). "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system." *Automation in Construction*, 19(5), 630–640.

Automated Information Retrieval for Hazard Identification in Construction Site

H. Kim[1], H. S. Lee[2], M. Park[3], B. Choi[4]

[1] Ph.D. Student, Dept. of Architecture and Architectural Engineering, Seoul National Univ., 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. PH +82-2-880-8311; FAX +82-2-887-8923; email: verserk13@naver.com
[2] Prof., Dept. of Architecture and Architectural Engineering, Seoul National Univ., 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. PH +82-2-880-8311; FAX +82-2-887-8923; email: hyunslee@snu.ac.kr
[3] Prof., Dept. of Architecture and Architectural Engineering, Seoul National Univ., 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. PH +82-2-880-8311; FAX +82-2-887-8923; email: mspark@snu.ac.kr
[4] M.S Course Student, Dept. of Architecture and Architectural Engineering, Seoul National Univ., 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Republic of Korea. PH +82-2-880-8311; FAX +82-2-887-8923; email: mill45@snu.ac.kr

ABSTRACT

The repetitive occurrence of similar accident in construction disasters is one of the prevalent features. Similar accident cases provide direct information for determining the risk of scheduled activities and planning safety countermeasure. Researchers have developed many systems in order to retrieve and use past accident cases. Although the developed systems have a clear and limited target, most of them were developed under a retrieval methods based on ad-hoc systems which can cause inconvenience for users in using the retrieval system. To overcome these limitations, this study proposes an automated information retrieval system that can search for and provide similar accident cases. The retrieval system extracts building information modeling objects and composes a query set by combining BIM objects with a project management information system. Based on the results of this study, the users can excessively reduce query generation. Furthermore, they can easily avoid risks by receiving similar past accident cases that can happen while they work.

INTRODUCTION

Despite the enormous efforts made in regards to safety, the construction industry has had a poor record of preventing accidents. One of the noticeable characteristics of construction accidents is repetitive occurrences (Abudayyeh et al. 2003). Accident cases are the strongest stimuli in safety management and raising laborer's awareness for safety. This is due to the knowledge from past accidents being directly related to the prevention of future accidents and laborers' raised safety awareness (Lindberg et al. 2010).

Information retrieval technology can be used to improve the efficiency and quality of safety management. Some approaches, such as case-based reasoning (CBR) and information retrieval systems based on accident cases, have attempted to put accident cases for practical use (Ye 1998). Despite the contributions of current systems to improve safety management, laborers have not adopted these systems because there might be avoidance and inconveniences in performing a large amount of queries to search for the desired information.

Therefore, this study suggests an information retrieval system that can overcome these limitations. The suggested system can automatically search for similar past accident cases. Based on the push system concept, the suggested system provides past accident cases to the laborers who are similarly related to the case.

BACKGROUND

User and Usability. Many researchers use safety information by relating construction activities or retrieving accident cases applied to safety planning. Several research studies about safety information systems have been performed at the national level. They provide the information on the related past accident cases when a user inputs queries. However, generating a query every time for a user to retrieve information makes the current systems inconvenient.

Since laborers are overconfident in their skills and experiences, they usually do not recognize the necessity of using past accident cases. This issue can be solved by the push system. The push system can be defined as automatically providing the right knowledge to the right person (Meso and Smith 2000). To provide similar accident cases to a laborer by using a push system, it is required to have the information about where an activity is performed and who is related to the activity.

Query generation and providing results. Generally, information retrieval can be achieved by using either Ad hoc systems or filtering systems (Jukna 2012). Ad hoc systems input queries whenever a user requires new information. Filtering systems store user interests in the user profile, and its main objective is to select newly generated information and delete useless information. Also, a routing system—which is one of the information filtering system—automatically pushes search results to users. If users want to identify countermeasures by retrieving accident cases, current safety retrieval systems using the Ad hoc model should input as many accurate queries as possible with the number of risk factors. However, with the routing model, users do not need to generate queries every time they retrieve. This model retrieves information whenever there is a change in the user profile or the DB. Also, it can push results to the related users.

Retrieval Method. Most of safety information retrieval systems have been developed using the Boolean model to compare queries and the DB, and to extract related cases. The Boolean model is processed based on whether or not the documents contain the query terms. Although the Boolean model is simple and clear, too many or too few results can be retrieved by its matching rule (Yves and Hammer 2010).

Another problem is that most researchers set the weight of each index equally. This leads to a decrease in the effectiveness and usefulness of retrieval results (Clote and Kranakis 2002). Lin and Soibelman (2009) suggested an extended Boolean model that is a hybrid retrieval type of a vector model which can overcome the limitation of the original Boolean model.

To achieve system improvement, this study suggests an information retrieval system that combines the routing model and the extended Boolean model. If these two models are combined, the system can automatically provide retrieval results including their ranks.

INFORMATION RETRIEVAL SYSTEM

Indices for information retrieval. In the previous studies (Lee et al. 2012a; Choudhry et al. 2009), various risk factors are extracted and evaluated. This study adopted the ideas of hazard identification and risk assessment performed in Lee et al.'s study. Therefore, this study adopts ten factors extracted from the study completed by Lee et al. (2012a). The extraction process consists of two steps. The first step is to collect various risk influence factors from research studies and disaster investigation on the code standard. By reviewing related studies, 27 factors were finally selected. The second step is to select factors with much more influence than others by reviewing the result from the survey conducted with 42 qualified safety managers.

Lee et al. (2012a) suggested influence factors affecting the occurrence of an accident and ten factors were selected by experts in safety management experience. Although all the factors are used as indices for composing queries, not all of them are used in every phase of safety planning. Safety management plans are usually organized into three steps: preliminary plan, monthly plan, and daily plan. The uncertainty of construction projects makes it necessary to acquire information at each safety planning phase. Therefore, classifying factors is required to make the fittest queries for each phase. The committee—which is composed of thirteen members, each with more than ten years of experience—had presented their opinions to determine the classification of the ten factors.

Also, determining the weights of indices is necessary to improve the effectiveness of the similarity measurement (Kolodner 1992). To weigh the indices, the analytic hierarchy process (AHP) developed by Saaty (1980) was used. The questionnaire for calculating weights follows that of Lee et al. (2012a)'s. It was sent to fifty experts, each of whom had worked more than ten years as a safety manager. The results of the surveys and AHP analysis are shown in Table 1.

To judge the degree of congruency between each index, it is important to determine the proper data format (Ye 1998). Depending on the type of data, the similarity index (SI) calculation method varies. If data follows a numeric format, the distance between two values that correspond to the construction site condition value and the accident case is calculated using Formula 1:

$$SI = 1 - \frac{|A - B|}{A} \quad (1)$$

where A = condition value of a current site; and B = condition value of a past accident case.

Table 1. Classification and weight indices by safety planning phase.

Influence Factors	Preliminary	Monthly	Daily
Work process rate	0.241	0.167	0.099
Cost of construction	0.175	0.131	0.077
Work type	0.334	0.238	0.125
Building type	0.250	0.173	0.096
Occupation type	-	0.181	0.100
Date	-	0.109	0.059
Age	-	-	0.069
Workdays on current site	-	-	0.118
Safety training	-	-	0.166
Temperature	-	-	0.089
No. of distributed surveys	50	50	50
No. of collected surveys	43	43	43
No. of collected surveys with a consistency index below 0.1	43	31	23

When the type of data follows a string format, SI is 1 if the condition value of a current site and the condition value of a past accident case are identical and 0 if they are completely different.

After calculating the similarity index of each influence factor, the similarity score (SS) was determined. This value can be expressed as the sum of the multiplication of the similarity index (SI) and weight (M), as seen in Formula 2:

$$ss = \sum_{i=1}^{n} M_i SI_j \qquad (2)$$

Data source for system development. Appropriate data source for information retrieval system is one of the most important aspects to develop the system. The structure of the data source in this study incorporates the benefits of PMIS and commercial BIM programs. Lee et al. (2012b) demonstrated the possibility of combining and linking information between PMIS and influence factors. Navon and Kolton (2006) suggested a model combining schedule and safety information based on AutoCAD or BIM. Based on these methods, the DB is established on the outside of a BIM program, and the values of influence factors and accident cases will be saved on to the DB. Using the extraction function of BIM, geometric values are extracted and connected with the DB. Through this kind of system composition, users can save, handle, and use information as much as they want. Also, property values can be used without decreasing the efficiency of commercial BIM programs.

Information retrieval system framework. The retrieval system framework suggested in this study is shown in Figure 1. The system consists of input, data processing, and output. The input has information about factors extracted from indices for the information retrieval section, weight of each factor, and accident case

DB. The weight of each factor's DB has weight values acquired through AHP analysis. These values are used as indices of weight in an information retrieval module. The accident case DB is utilized to search similar accident cases by comparing queries based on construction site conditions.
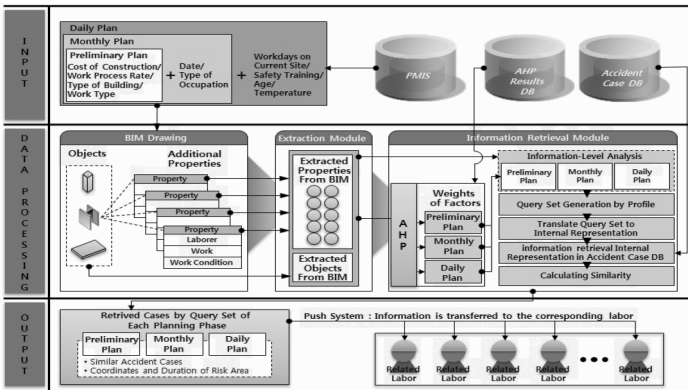

Figure 1. Retrieval system framework.

Data processing comprises the BIM drawings module, extraction module, and information retrieval module. The BIM drawings module forms 3D oriented parametric models. It has the geometric property values of objects, and provides objects that include additional property values. The extraction module extracts objects from BIM drawings. Finally, extracted objects and information corresponding with influence factors are combined. Information processed in the extraction module is sent to the information retrieval module and is used to find similar accident cases.

The most relevant case for each labor can be found by this type of query generation. PMIS including information related to laborers is combined with BIM objects. This combination makes a query set. In this process, property values of a labor category generate a suitable query for each laborer. For example, the crews that work on the same activity have the same values of work and work condition. However, an individual labor's property values vary. These property values are incorporated to the generating query process. The generated queries include not only characteristics of work and work condition but also characteristics of individual labor.

The information retrieval module determines the type of retrieval method by the number and degree of BIM property values. After that, AHP result values are loaded by the type of retrieval method. Also, queries are generated and converted to internal representations. Based on query sets by the safety planning phase, the system performs retrieval of similar past accident cases. Similarity indices are calculated by comparing accident cases and internal representations based on construction site values.

The output represents similar past accident cases retrieved by query sets of each safety planning phase. The similar accident cases classified by the number of inputted factors can be found. Each case provides the coordination and remaining

period of risk factors by combining objects extracted from BIM drawings. Retrieved past accident cases are extracted by the property values connected with a BIM object. The object connected with retrieved past accident cases includes information consisting of labor work type, work start time, work finish time, coordinates of work area, and etc. as property values. The suggested system searches for the laborers whose work type or work conditions correspond to the retrieved past accident cases. Then, these cases will be provided to related labors. The BIM object including geometric information and additional properties has an identifier. Thus, the labor ID in additional properties can be linked with an object's identifier. The identifier is used to search a labor / laborers who has/have relevance to the object. The system provides retrieval results to related laborers.

APPLICATIONS

The sample case is an apartment building construction project located in Seoul, Korea. This construction project had been operated from July 2009 to April 2012. The project consists of 4 buildings with 24 stories and 340 households. The suggested system had been used for this project in 2 months (February to March 2011).

The results of information retrieval are shown in Figure 2. Safety managers and laborers were provided with similar past accident cases sorted by similarity scores. Query was generated by the combination of influence factors. Based on the query set, the retrieval process was performed and the results presented the similarity of each past accident case, as well as the related laborers' information. The cases included the similarity score, the related laborers, coordinates of the risk area extracted from BIM objects, and the duration of risk factors.
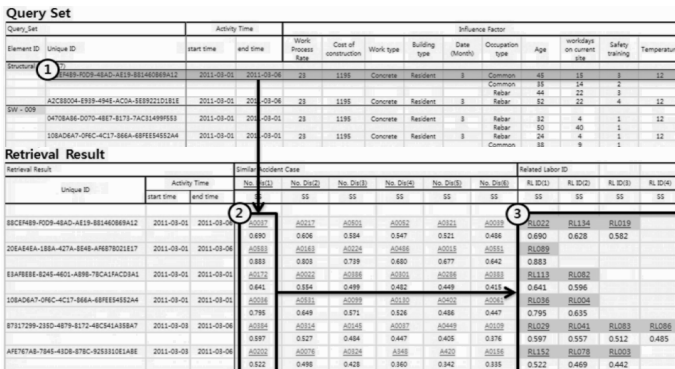

Figure 2. Example of the information retrieval results.

The suggested system was validated by comparison the retrieval performance of the system with the established accident case retrieval system of KOSHA. Precision and recall are the two most frequent and basic measurement indices for information retrieval effectiveness (Manning et al. 2008). The definitions of precision and recall are as follows:

$$\text{Precision} = \frac{\#(relevant\ items\ retrieved)}{\#(retrieved\ items)} = P(relevant\ /\ retrieved)\ (3)$$

$$\text{Recall} = \frac{\#(relevant\ items\ retrieved)}{\#(relevant\ items)} = P(retrieved\ /\ relevnat)\ (4)$$

Generally, text retrieval conference (TREC) collections have been used to evaluate the performance of an information retrieval system (Zobel et al. 1996). However, it is difficult for the suggested system to directly apply the text collections of TREC because the DB is structured based on past accident cases. Therefore, the precision and recall of this study were evaluated through the comparison of KOSHA's past fatal accident case retrieval system. One hundred past accident cases were extracted from KOSHA's fatal construction accident DB by using the random extraction function. The calculations of precision and recall were performed by reflecting safety steps: preliminary plan, monthly plan, and daily plan. The number of indices of each safety plan is different. Precision and recall were estimated by using Formulas 3 and 4, and the results are presented as Table 2.

Examining the results of measurements, the values of precision and recall were higher than those of KOSHA's retrieval system. This indicates that the suggested system has the potential to retrieve more related information, and the retrieved cases can be more likely to be used. Moreover, the retrieval performance of the suggested system can be considered higher than that of the currently used system.

Table 2. Retrieval effectiveness comparison results.

Safety planning phase	Preliminary plan		Monthly plan		Daily plan	
Number of indices	4		6		10	
Search system	P	R	P	R	P	R
KOSHA system	0.72	0.61	0.63	0.41	0.84	0.34
Suggested system	0.68	0.82	0.79	0.75	0.92	0.61

CONCLUSION

This study proposes an accident case retrieval system that can automatically generate queries based on construction site conditions. The results include time and geometric information, as well. Also, they are automatically provided to the laborers by using the push system. To develop the suggested information retrieval system, BIM, PMIS, AHP result DB, and the accident case DB were structured. Then, information retrieval algorithms were defined. Finally, the push system was established to provide retrieved accident cases to the related laborers. The results of the suggested system include similar cases happened in the past, related works, and the coordination of similar cases and work time in a current construction site. These results can help safety managers prepare safety countermeasures in the safety planning steps and also help raise laborers' attention to safety. The automatic retrieval of accident cases in the system increases the efficiency of identifying risk factors.

Despite these advantages, this study has some limitations. If there is no abundant data source such as PMIS, enormous efforts are required to input data. Also,

accumulated accident cases may not be sufficient for application to safety management. Interworking between 3D models and the coordination of the retrieval result is needed to visualize the risk area or factors. Future research will eventually remedy these limitations and help improve safety management.

ACKNOWLEDGEMENT

REFERENCES

Abudayyeh, O., Federricks. T., Palmquist, M., and Torres, H. (2003) "Analysis of Occupational Injuries and Fatalities in Electrical Contracting." J. Constr. Eng. Manage., 129(2), 152-158.

Clote, P. and Kranakis, E. (2002). Boolean Functions and Computation Models, Springer, New York, USA.

Jukna, S. (2012). Boolean Function Complexity: Advances and Frontiers. Springer, New York, USA.

Kolodner, J. L. (1992). "An introduction to case-based reasoning." Artificial Intelligence Review, 6(1), 3-34.

Lee, H., Kim, H., Park, M., Teo, E. A. L., and Lee, K. (2012a). "Construction risk assessment using site influence factors." J. Comput. Civ. Eng., 26(3), 319-330.

Lee, H., Lee, K., Park, M., Baek, Y., and Lee, S. (2012b). "RFID-Based Real-Time Locating System for Construction Safety Management." J. Comput. Civ. Eng., 26(3), 366-377.

Lin, K. and Soibelman, L. (2009). "Incorporating domain knowledge and information retrieval techniques to develop and architectural/engineering/construction online product search." J. Comput. Civ. Eng., 23(4), 201-210.

Manning, C. D., Raghavan, P. and Schutze, H. (2008). Introduction to information retrieval, Cambridge University Press, New York, USA

Meso, P. and Smith, R. (2000). "A resource-based view of organizational knowledge management systems." Journal of Knowledge Management, 4(3), 224-234.

Navon, R., and Kolton, O. (2006). "Model for Automated Monitoring of Fall Hazards in Building Construction." J. Constr. Eng. Manage., 132(7), 733-740.

Satty, T. L. (1980). The analytic hierarchy process, McGraw-Hill, New York. USA.

Ye, T. (1998). "Reasoning Model of the Case-Based Construction Safety Management System." Master Course. dissertation, Seoul National Univ., Seoul, Korea.

Yves, C. and Hammer, P. L. (2010). Boolean Models and Methods in Mathematics, Computer Science, and Engineering, Cambridge University Press, London.

Zobel, J., Moffat, A, and Ramamohanarao, K. (1996). "Guidelines for presentation and comparison of indexing techniques." ACM SIGMOD Record, 25(3), 10-15.