Marc Garbey • Barbara Lee Bass
Christophe Collet • Michel de Mathelin
Roger Tran-Son-Tay

Computational Surgery and Dual Training

Springer

Computational Surgery and Dual Training

Marc Garbey • Barbara Lee Bass
Christophe Collet • Michel de Mathelin
Roger Tran-Son-Tay
Editors

Computational Surgery and Dual Training

Springer

Editors
Marc Garbey
University of Houston
Houston, TX, USA
garbey@cs.uh.edu

Barbara Lee Bass
The Methodist Hospital Research Institute
Houston, TX, USA
BBass@tmhs.org

Christophe Collet
University of Strasbourg
Illkirch, France
c.collet@unistra.fr

Michel de Mathelin
University of Strasbourg
Illkirch, France
demath@eavr.u-strasbg.fr

Roger Tran-Son-Tay
University of Florida
Gainesville, FL, USA
rtst@ufl.edu

ISBN 978-1-4419-1122-3 e-ISBN 978-1-4419-1123-0
DOI 10.1007/978-1-4419-1123-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2009942085

This book is dedicated to the memory of Professor Michel Israël, "Conseiller pour la science et la technologie" from the French Embassy in Washington, DC.

Preface

The future of surgery is intrinsically linked to the future of computational sciences: the medical act will be computer assisted at every single step, from planning to post-surgery recovery and through the surgical procedure itself.

Looking back at the history of surgery, surgery practice has changed dramatically with the extensive use of revolutionary techniques, such as medical imaging, laparoscopy, endoscopy, sensors and actuators, and robots. This trend is dependent on the use of computer processing, computational method, and virtualization.

Computational surgery will not only improve the efficiency and quality of surgery, but will also give new access to very complex operations that require extreme precision and minimum intrusion. Such examples are today's inoperable cancer tumors that have invaded critical tissues or nervous centers. In order for this milestone to be reached quicker and more efficiently, surgeons will have to become very familiar with computing methods, such as image analysis, augmented reality, and/or robotics. It will be critical for surgeons to assimilate computers in their training, understand how computers work, understand the limitations/advantages of these computer tools, and be able to interpret computer imaging and simulations.

The goal of this book is to provide the background on and examples of computational surgery and training in that emerging field. This project started with the workshop "Computational Surgery and Dual Training" that was organized in Strasbourg, France in December 2008, which brought together a group of computer scientists, engineers, and surgeons to discuss interventional procedures and surgeries. The material that follows is organized into five parts which include 17 chapters that take the reader from the day-to-day practice of surgeons to computational optimized surgery methods that are in the making, through disease management, diagnostic, intervention, and the use of modeling and simulation. The last part deals with the fundamental question of training surgeons and how to prepare for the future of surgery while balancing complex objectives, such as cost effectiveness, risk management, and quality of life.

We present first the management of diseases in surgery practice. The domain of surgery is vast and divided into many disciplines from cardiovascular surgery, digestive surgery, neurosurgery, and cancer surgery, involving every system in the human body. We will proceed with representative examples that are well understood by the community for which recent progress in technology has brought changes

in clinical management. In breast cancer, for example, significant improvement in survival is attributed to a combination of factors that include improved screening with medical imaging, chemotherapy, radiotherapy, and surgery. However, quality of life depends not only on cancer cure, but satisfaction with the cosmetic results of breast-conserving and reconstruction therapies. We will then present a classic family of problems in endovascular surgery for lower extremity arteries. Endovascular surgery is a relatively new field of surgery that has seen exponential growth with the introduction of minimally invasive procedures, such as endovascular stenting, which have allowed outpatient care. At the same time, there might not be enough insight to carefully assess this new technology. This field of surgery provides a good example of how computational methods, starting from imaging to virtualization through hemodynamic simulation, may improve the decision-making and document efficiency and risk management. However, there is no shortage of computerized information in surgery: doctors acquire large data sets of imaging and all kinds of medical recordings before, during, and after surgery. This overwhelming stream of information can be challenging to synthesize. It is becoming critical indeed to build human computer interfaces that fit the surgeon's needs.

Part I Computer Assisted management of Disease and Surgery is divided into two chapters: Chap. 1 "Breast-Conserving Therapy for Breast Cancer: Targets for Investigation to Improve Results" by Bass and Garbey and Chap. 2 "Changing Paradigms in the Management of Peripheral Vascular Disease – The Need for Integration of Knowledge, Imaging, and Therapeutics" by Davies and Vykoukal.

The work of surgeon does not start in the surgery room and end there. The surgeon must work collaboratively with the radiologist at every step of the diagnostic process.

Medical imaging has been characterized by an explosion of image modalities that combine various forms of ultrasound, X-ray, MRI, PETS, etc… and protocols. It is a difficult exercise to read a medical image and use it for a diagnostic that goes far beyond anatomy. Many images are indirect measurement of metabolic activity or various form of "functioning." Imaging is quite specialized indeed and can come in different forms depending on the organ or simply the scale of interests. Presented are imaging techniques for the brain that have been specifically developed to identify lesion location and extend. Every brain is different indeed so pattern recognition, mapping, and classification, which are natural tools to support the medical decision, are also discussed.

Moving to vascular disease, we will discuss computational methods to identify vessels. These methods are fundamentally related to blood flow visualization. Discussed are imaging analysis techniques for cardiovascular disease and how they can be used by surgeons in their diagnostics. A particular property of blood that is often overlooked is that blood flow carries thermal energy all over the human body. Thermal imaging is a passive sensor that relies heavily on this energy that can be easily detected in vessels proximal to the skin. While the promise of 1980s thermal imaging did not live up to its expectations because of the limitation of equipment and lack of rigor in the exploitation images, there is currently a renewed interest in thermal imaging technology that can be brought to the surgery room. Discussed are

some of these developments in the identification of cardiac pulse from a distance, diagnostic tools and their ability to handle qualitative information at the cell level, and image processing. We will also discuss how high-performance computing can perform pattern identification for fast diagnostics.

Part II Image Processing and Diagnostics is divided into five chapters; Chap. 3 "Brain MRI Segmentation" by Bricq et al., Chap. 4 "Knowledge-Driven Recognition and Segmentation of Internal Brain Structures in 3D MRI" by Bloch and material on Cardio including Chap. 5 "New Dimensions in Diagnostic Imaging of the Aorta" by Bismuth et al.; Chap. 6 "Methodological Advances on Pulse Measurement Through Functional Imaging" by Bourlai and Pavlidis, and on cells Chap. 7 "Parallel Multispectral Image Segmentation for Computer Aided Thyroid Cytology" by Shah and Gabriel.

Medical imaging has been used primarily as pre-processing or post-processing steps. Such steps include diagnostic, surgery planning, and recovery assessment. However, minimal invasive surgery cannot proceed without bringing a camera into the instrumentation. Consequently, surgeons end up operating through the eyes of technologic tools: their eyes are not on the patient but instead focus on the display monitor. For instance in the case of abdominal intervention, the surgeon's hands are extended by laparoscopy tools with less than normal, natural control. Eventually, for zones in the body with difficult access, the surgeon's hands act on these parts indirectly through the arms of a robot. Computers improve by de facto the way the surgeon works, by relaying their actions through complex computing procedure to reach better accuracy, minimize damage to surrounding tissues, and deliver the right procedure at the right location.

Part III Image Driven Intervention and Robotics describes state of the art procedures in digestive surgery to vascular surgery, MRI guided intervention in surrounding tissues. The discussion begins with Chap. 8 "Computer-Assisted Digestive Surgery" by Soler et al., followed by Chap. 9 "Design of a Robotized Flexible Endoscope for Natural Orifice Transluminal Endoscopic Surgery" by de Mathelin followed by Chap. 10 "MRI-Guided Robot-Assisted Interventions: An Opportunity and a Challenge in Computational Surgery" by Tsekos et al., and concludes with Chap. 11 "Image-Guided Interventions and Robotics" by Bayle et al.

Computational surgery has been a step by step incremental improvement of medical imaging. One starts with the collected image, analyzes it by computational procedures, and runs this process interactively during the course of the surgery. This process provides a new visual tool through dynamic imaging processing that can bring to light information that would otherwise be out of the reach of the surgeon. This in return impacts how the surgeon will practice. Modeling and Simulation have more to do with how things work, and how one can evaluate the effects of surgery. The goal of modeling and simulation in computational surgery is to anticipate based on how the system works. Hemodynamics is more complex than classical fluid dynamics: blood vessels are living organisms with viscoelastic properties that feel the flow and react to it. The time scale involved goes from 1 Hz which is basically the heart beat frequency to dozen of years which is the scale of atherosclerosis build up. Presented is a system biology approach of the vein graft

intervention that is a rare example of a practical attempt to integrate multiscale components of a cardiovascular disease into an integrated computational model of the disease. Fundamental mechanisms at the cell level that address the inflammatory process which is key in many diseases will be given to complete the picture. Macroscopic evidence of how fluid dynamic properties relate to hemodynamic diseases with practical clinical cases is given to address the relevance of simulation in endovascular surgery.

A similar approach can be taken when cancer is the diagnosis. We will present a macroscopic model approach for breast cancer that can be patient-specific and addresses some of the clinical issues.

Part IV Modeling, Simulation and Experimental data starts with Chap. 12 "Emerging Mechanisms of Vein Graft Failure: The Dynamic Interaction of Hemodynamics and the Vascular Response to Injury" by Berceli et al., and is followed by Chap. 13 "Modeling and Role of Leukocytes in Inflammation" by Tran-Son-Tay et al. and progresses to Chap. 14 "Multi-Modality Imaging for the Simulation of Cerebral Aneurysm Blood Flow Dynamics" by Karmonik et al., and concludes with Chap. 15 "A Computational Framework for Breast Surgery: Application to Breast-Conserving Therapy" by Thanoon et al.

Computing has infiltrated every step of the surgeon's work, changing the practice to the point that what was learned by a surgeon 15 years ago during formal training has very little to do with practice in the surgery room today. The technological sophistication is evolving so fast that training must adapt as well. To determine what training is needed, a rigorous assessment of the responsibilities and activities of practicing surgeons should be done. Clearly, continuing education and computer simulation will play an important role. Presented here for the first time, is an integrated system that streamlines the work of a surgeon and supports the decision-making process with the use of a standalone computational environment that can also be used for training in computational surgery.

Part V, Training presents Chap. 16 "Simulators in Training" by Dunkin and Chap. 17 "A Computational Desk for Surgeons" by Hilford et al.

Computational surgery is still in its infancy and training in the field is scarce. Although the material presented here on computational surgery is not comprehensive, it nevertheless fills a necessary gap. The goal of this monograph is to provide the necessary background and examples to medical doctors and scientists so that they can speak the same language and communicate more effectively. We hope that, by selecting these different topics, we have provided a tool for training in and a better appreciation of the field of computational surgery. We refer to the Web site of computational surgery.org that will keep this effort growing. Finally, we would like to thank the Partner University Fund (PUF) that supported this team project.

Houston, TX — *Marc Garbey*
Houston, TX — *Barbara Lee Bass*
Illkirch, France — *Christophe Collet*
Illkirch, France — *Michel de Mathelin*
Gainesville, FL — *Roger Tran-Son-Tay*

Contents

Contributors

Vincent Agnus, PhD IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Jean-Paul Armspach, PhD University of Strasbourg, Strasbourg, CNRS, France, jparmspach@unistra.fr

L. Barbé, PhD LSIIT, University of Strasbourg, France,

Bérengère Bardou LSIIT Laboratory, University of Strasbourg, CNRS, France, bardou.berengere@unistra.fr

Barbara L. Bass, MD Department of Surgery, The Methodist Hospital, 7111 Fannin St, Houston, TX, USA
and
Weill Medical College of Cornell University, New York, NY, USA, BBass@tmhs.org

B. Bayle, PhD LSIIT, University of Strasbourg, Strasbourg, France, bernard.bayle@unistra.fr

Scott A. Berceli, MD, PhD University of Florida, Box 100128, Gainesville, FL, 32610, USA
and
Department of Surgery, University of Florida College of Medicine, Gainesville, FL, USA, Malcom Randall Veterans Affairs Medical Center, Gainesville, FL, USA, bercesa@surgery.ufl.edu

Jean Bismuth, MD The Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA, JBismuth@tmhs.org

Isabelle Bloch Telecom ParisTech, CNRS UMR 5141 LTCI, Paris, France, isabelle.bloch@enst.fr

Thirimachos Bourlai, PhD Computational Physiology Lab, University of Houston, 4800 Calhoun Rd, Houston, TX, USA, tbourlai@uh.edu

Stéphanie Bricq, PhD LSIIT Laboratory, University of Strasbourg, Strasbourg, CNRS, France, bricq@lsiit.u-strasbg.fr

Pradeep Buddharaju Computational Physiology Lab, University of Houston, 4800 Calhoun Rd, Houston, TX, USA, pbuddharaju@uh.edu

Arnaud Charnoz, PhD IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Christophe Collet, PhD LSIIT Laboratory, University of Strasbourg, Strasbourg, CNRS, France, c.collet@unistra.fr

Bernard Dallemagne IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Mark G. Davies, MD, PhD, MBA Methodist DeBakey Heart and Vascular Center, Department of Cardiovascular Surgery, The Methodist Hospital, Houston, TX, USA, MDavies@tmhs.org

Michel de Mathelin, PhD LSIIT laboratory, University of Strasbourg, Strasbourg, CNRS, France, demath@eavr.u-strasbg.fr

Brian J. Dunkin, MD Department of Surgery, Weill Cornell College of Medicine, The Methodist Hospital, Houston, TX, USA, bidunkin@tmhs.org

Jean-Baptiste Fasquel IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Edgar Gabriel, PhD Department of Computer Science, University of Houston, Houston, TX 77204, USA, gabriel@cs.uh.edu

Marc Garbey, PhD Department of Computer Science, University of Houston, 501 Phillip Hoffman Hall, Houston, TX 77204, USA, garbey@cs.uh.edu

Robert G. Grossman The Methodist Hospital, Neurological Institute, Houston, TX, USA

Victoria Hilford University of Houston, Houston, TX, USA, vhilford@cs.uh.edu

Alexandre Hostettler IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Minki Hwang Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA

Christof Karmonik, PhD The Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA
and
The Methodist Hospital Neurological Institute, Houston, TX 77030, USA, CKarmonik@tmhs.org

Nam-Ho Kim, PhD Department of Mechanic and Aerospace Engineering, University of Florida, Gainesville, FL, USA

Alan B. Lumsden, MD The Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA, ablumsden@tmhs.org

Jacques Marescaux, MD IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Johan Moreau, PhD IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Didier Mutter IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Florent Nageotte LSIIT laboratory, University of Strasbourg, CNRS, Strasbourg, France

Stéphane Nicolau IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France

Ioannis Pavlidis, PhD Computational Physiology Lab, University of Houston, 4800 Calhoun Rd, Houston, TX, USA, ipavlidis@uh.edu

O. Piccin LSIIT, University of Strasbourg, Strasbourg, France,

P. Renaud LSIIT, University of Strasbourg, Strasbourg, France,

Dipan Shah The Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA

Shishir Shah, PhD Department of Computer Science, University of Houston, Houston, TX 77204, USA, sshah@central.uh.edu

Luc Soler, PhD IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France, luc.soler@ircad.fr

Ahmet Eren Sonmez Medical Robotics Laboratory, Department of Computer Science, 4800 Calhoun – 501 Phillip G. Hoffman, University of Houston, Houston, TX 77204-3010, USA

David Thanoon Department of Computer Science, University of Houston, 501 Phillip Hoffman Hall, Houston, TX 77204, USA, dthanoon@cs.uh.edu

Roger Tran-Son-Tay, PhD Department of Mechanical and Aerospace Engineering, University of Florida, Box 116250, Gainesville, FL 32611, USA
and
Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA, rtst@ufl.edu

Nikolaos V. Tsekos, PhD Medical Robotics Laboratory, Department of Computer Science, 4800 Calhoun – 501 Phillip G. Hoffman, University of Houston, Houston, TX 77204-3010, USA, nvtsekos@central.uh.edu

Daynene Vykoukal, PhD The Methodist Hospital Research Institute, Houston, TX, USA, DVykoukal@tmhs.org

Erol Yeniaras Medical Robotics Laboratory, Department of Computer Science, 4800 Calhoun – 501 Phillip G. Hoffman, University of Houston, Houston, TX 77204-3010, USA

Yusuf Yildiz University of Houston, Houston, TX, USA, yildizyusuf@gmail.com

Philippe Zanne LSIIT laboratory, University of Strasbourg, CNRS, France

Yi Jonathan Zhang, MD The Methodist Hospital Neurological Institute, Houston, TX, USA

Part I
Computer Assisted Management of Disease and Surgery

Breast-Conserving Therapy for Breast Cancer: Targets for Investigation to Improve Results

Barbara Lee Bass and Marc Garbey

Breast Cancer Biology

Molecular characteristics of breast tumors are linked to their clinical behavior, with estrogen and progesterone receptor (ER/PR) expressing tumors, showing a less aggressive pattern of growth and metastasis, while ER/PR negative tumors behave in a more aggressive biologic fashion [1]. Expression of the EGF receptor-like moiety HER2-neu is also associated with adverse clinical behavior. The definition of tumor molecular signatures and clinical phenotype is undergoing redefinition with the description of luminal and basal-like categories [2]. Characteristic clinical behaviors and responsiveness to hormonal and chemotherapeutic definitions is predictable based on these features and continued definition of molecular targets for therapy is an important ongoing area of investigation. As the molecular pathways that govern tumor initiation, growth and metastasis are defined, it is certain that new effective targeted therapies will be developed to fundamentally direct our methods for treating breast cancer.

At the present time, the management of breast cancer is based on the coordinated therapeutic modalities of surgery, anti-tumor medications, and radiation therapy. The goals of each modality are based on the current biological hypothesis of breast cancer as a systemic disease. This hypothesis states that breast cancer is a systemic disease at the time of tumor initiation. It is presumed that circulating tumor cells are shed by the primary tumor, modulated by specific molecular characteristics of the malignancy, from the time when the tumor first develops [3]. Initially cleared by immunologic mechanisms, occult systemic metastases develop by entrapment and implantation of circulating tumor cells in remote organs. Cellular features of the tumor, including high cytological grade and rapid proliferative rate, are predictive

B.L. Bass (✉)
The Methodist Hospital Research Institute, Houston, TX, USA
e-mail: BBass@tmhs.org

M. Garbey
University of Houston, TX, USA
e-mail: garbey@cs.uh.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_1,

of greater risk for metastasis, but the most reliable predictor of metastatic disease is the amount of tumor that has spread to the primary regional lymph node basin of the breast, the axillary nodal basin [4]. Breast cancer is graded by stage based on the American Joint Commission on Cancer staging system (AJCC) [5]. The disease of each patient with breast cancer is graded according to the tumor size (T), the presence and degree of node metastasis (N), and the presence or absence of detectable metastases (M). The TNM grades are then summated to establish a Stage for each patient with breast cancer. Stage I and II are considered early stage breast cancers, with 5-year survival rates of 98 and 85% respectively, while Stage III disease carries a 5-year survival rate of 55–65%. Stage IV is characterized by the presence of metastatic disease and survival rates are 20–30% at 5 years. Hence early detection and effective treatment of early stage breast cancer offers excellent opportunities for long-term survival and "cure."

On the basis of biological hypotheses of breast cancer as a systemic disease at presentation, the goals of each modality of therapy for Stage I–III breast cancer is clear. First, the role of surgery is to fully remove the primary tumor in the breast to achieve local tumor control. Second, surgery is used to establish the presence and degree of tumor burden in the axillary nodal basin as a guide for adjuvant (medical) therapies, and if nodal metastases are present, to remove the axillary lymph nodes to minimize the risk of locoregional recurrence. The goal of radiation therapy is to optimize locoregional control of the tumor by treating the at-risk region, the residual breast, or the chest and axillary region, with ionizing radiation to selectively destroy presumed microscopic residual tumor. In Stage I–III breast cancer patients, in whom there is no detectable disease, hormonally-based, molecularly targeted or cytotoxic chemotherapy, medications are administered to eradicate or control presumed occult residual metastatic tumors [6,7]. The use of chemotherapy in patients with presumptive metastases based on known risk for subsequent recurrence of cancer based on the Stage of presentation, but in whom disease is not detectable by any objective measure such as imaging or blood analysis, is known as adjuvant therapy. An inherent principle of adjuvant therapy is that some patients will be "over treated," i.e., the individual patient may not actually have any residual micrometastases at risk for subsequent growth, but as a member of a population at risk based on tumor stage, the patient is treated presumptively. Future work to better identify micrometastatic disease based on refined molecular profiling of primary tumors, development of more sophisticated imaging modalities to identify micrometastatic deposits, and other avenues of investigation, will hopefully offer more precise means to guide systemic therapy.

Surgical Management of Early Breast Cancer

Surgery remains an essential component of breast cancer care in all patients with Stage I–III breast cancer. The objective of surgery is to fully remove the tumor to minimize the risk of disease recurrence in the breast or chest region, and to optimize

the cosmetic result to enhance the patient's sense of self and well-being. Two surgical resection options are available: total removal of the breast – mastectomy or breast conserving therapy coupled to radiation therapy of the residual breast, in which the primary tumor is removed from the breast leaving the remaining breast in place. Mastectomy is necessary for patients with multiple tumors within the breast, patients who are unable to complete post-operative radiation therapy, and for a variety of other conditions. Mastectomy is well-tolerated by many patients, however recovery may be complicated by wound disruption, impaired shoulder mobility, and patient anxiety and depression due to changes in body image and the ever present reminder of their cancer diagnosis evidenced by the absent breast. Reconstructive surgery is appropriate for many patients undergoing total mastectomy. Older patients often decline reconstruction due to the challenges associated with multiple surgical procedures. Younger women typically pursue some form of reconstructive surgery and in early stage cancer, these reconstructive procedures are performed at the same surgery as the mastectomy. Reconstructive options include the use of implant-based procedures to restore the breast mound and the use of more complex, autologous tissue transfers utilizing the abdominal wall, posterior chest, or gluteal soft tissues. The cosmetic results of these procedures show that most patients are highly satisfied with their reconstructed breast, although these procedures are physiologically challenging for many patients.

Fortunately, up to 70% of patients with the new diagnosis of breast cancer have tumors that are amenable to breast-conserving surgery. A national multi-site randomized controlled trial first reported in 1986, and now with follow-up intervals in excess of 20 years, has shown that breast-conserving therapy (BCT), complete removal of the tumor coupled to post-operative breast radiation therapy, is associated with equivalent survival compared to total mastectomy for the treatment of early stage breast cancer [4, 8, 9]. Breast-conserving surgery is a shorter, better tolerated, and simpler surgical procedure for most patients than total mastectomy. BCT is considered the preferred method of surgical management of breast cancer in patients who are candidates for this procedure.

Several principles must be observed in BCT. First, the entire tumor must be removed – a procedure known as lumpectomy or partial mastectomy – and all surgical margins must be microscopically free of tumor cells. Second, the patient must complete some form of radiotherapy after surgery to minimize the risk of local recurrence of tumor in the breast. Multiple modalities of radiotherapy are now available, although only whole breast external beam radiotherapy has been compared in a randomized clinical trial in comparison to total mastectomy. Whole breast radiotherapy, delivered at a total dose of 50 Gy with a targeted boost to the tumor bed at daily intervals over a 5–6-week period, can reduce the risk of local recurrence to 4–8%. Additional modalities that are demonstrating equivalent efficacy with greater efficiency and convenience for the patient in ongoing registry data include intraoperative radiotherapy, catheter-based brachytherapy methods, and accelerated partial breast external beam radiotherapy. Results of ongoing randomized clinical trials will define optimal strategies for radiotherapy following lumpectomy.

A third principle of breast-conserving surgery is that the cosmetic result should be satisfactory to the patient. Numerous factors contribute to the cosmetic result after lumpectomy, including the relative size of the tumor to the overall breast, the position of the tumor in the breast, the density of the patient's breast tissue, the impact of post-operative radiotherapy, or the occurrence of a post-operative infection. In patients with large breasts and small tumors, lumpectomy can often be accomplished with minimal breast deformity, whereas, women with smaller breasts and larger tumors may face considerable breast deformity with lumpectomy procedures. There is an inherent tension between wide excision of the tumor to achieve negative microscopic margins with the desire of the surgeon to remove only so much tissue as is necessary to optimize the cosmetic result of lumpectomy. This problem is compounded by the pattern of growth of most tumors in which microscopic stellate extensions into the surrounding breast tissue are not visible to the surgeon at the time of surgery. Furthermore, it is not feasible using current intraoperative histological techniques to adequately assess the margins of excision at the time of surgery in patients undergoing lumpectomy. Hence, up to 15–20% of patients may need to return for re-excision lumpectomy to achieve negative margins prior to proceeding on to radiotherapy. Re-excision, not only entails another operation for the patient but also is associated with less favorable cosmetic outcome [10].

For patients with a high tumor:breast volume ratio, the cosmetic result of breast-conserving therapy can be enhanced with the use of pre-operative hormonal or chemotherapy. Estrogen receptor expressing tumors may decrease in size substantially with a 3–4-month course of anti-estrogen therapy, either the selective estrogen receptor modulator tamoxifen or one of the aromatase inhibitors which interrupt the estrogen synthetic pathway. For patients in whom adjuvant cytotoxic systemic therapy will be indicated based on AJCC Stage at presentation, such as T2 or greater size or a clinically positive axillary nodal basin, neoadjuvant (therapy prior to surgery) chemotherapy may be administered with the goal of decreasing the size of the primary tumor, thereby decreasing the amount of breast tissue that will need to be removed to achieve negative margins during lumpectomy. The smaller the volume of tissue that needs to be removed relative to residual breast volume, the better the cosmetic outcome. Hence, neoadjuvant hormonal and cytotoxic approaches can be an aid in BCT.

Targets for Improvement in BCT Surgery

While BCT is the preferred therapy for most patients with early stage breast cancer, several aspects of the treatment remain suboptimal and are good targets for investigation using computational surgery methods. These elements are (1) the ability to predict the contour of the breast after lumpectomy, (2) the integration of breast

imaging into surgical planning, (3) the ability to predict the impact of radiation on breast healing and cosmesis, and (4) the ability to predict tumor extension and to accurately assess surgical negative margins during surgery.

The Impact of Surgical Lumpectomy on Breast Contour

Surgical lumpectomy is performed based on the surgeon's perceived placement of the tumor within the breast based on the physical examination and pre-operative breast imaging. Breast imaging modalities of mammography, ultrasonography and increasingly breast MRI are invaluable in surgical planning for lumpectomy. Each, however, provides a distinctly different perspective on the breast mound and while each contributes to surgical planning, none offers a comprehensive or synthesized image for surgical guidance. Mammography offers compressed two-dimensional views of the breast oriented in the vertical and horizontal planes. Mammography helps place the targeted lesion in a quadrant of the breast but does not offer specificity of site. Ultrasound provides non-registered views of lesions within the breast parenchyma that enhance the surgeon's ability to judge shape and size of the gross, but not microscopic tumor. MRI is the most sensitive imaging modality, and coupled to three-dimensional modeling software, provides the best pre-operative visualization of the tumor in the breast [11]. However, in current applications, MRI may over or under estimate the size of the breast tumor so surgeons cannot routinely use this modality to guide lumpectomy. None of these modalities, given their primary role as tissue imaging modalities, provides any measure of breast surface contour. Interestingly, however, it is the breast surface contour that is most central to cosmetic outcomes after lumpectomy.

To date, these three modalities do not allow co-registration of data points and so integration of these images becomes a cognitive function performed by the surgeon. With clinical experience, surgeons have developed surgical techniques to optimize cosmetic results; however, it is generally true the actual breast contour result after lumpectomy is not accurately predictable. We hypothesize that surgical planning can be improved and that breast contour could be better predicted using virtual lumpectomy modeling based on pre-operative imaging data.

In this monograph, Thanoon et al. reports on theoretical modeling of the impact of lumpectomy on breast contour. Our current studies are investigating the impact of virtual lumpectomy based on clinical data. A valuable clinical endpoint would be the ability to plan surgical lumpectomy based on preoperative imaging that would reliably predict the post-operative contour of the breast. Multiple clinical parameters besides simple change in tissue volume apply to this modeling. Variables include breast tissue density, the wound healing process, and preservation of the skin envelop. Breast density is highly variable in women. Premenopausal women have more glandular and dense fibrous tissue in the breast

while post-menopausal women generally have replacement of glandular elements with fatty and fibrous tissue elements. Some patients have extremely dense breast parenchyma associated with fibrous proliferation and minimal tissue deformability. These latter patients are at greater risk for breast deformity after lumpectomy. The degree of tissue compliance and deformability impacts on final breast contour. Tissue density can be qualitatively, but not quantitatively assessed on pre-operative mammography [12, 13].

The wound healing process is also highly dynamic after breast lumpectomy. Immediately after lumpectomy, a reactive inflammatory process characterized by development of a fluid-filled cavity (seroma) develops, often precisely filling the space once occupied with breast tissue. However, as the healing process advances, the seroma is reabsorbed and mesenchymal tissue elements of scar such as myofibroblasts occupy the space and induce contraction of the wound. The final surface contour is significantly influenced by this contracture process and must be considered in predictive models. Contracture of the wound can lead to considerable concave deformity of the breast.

The contour of the breast is confined by the skin envelop. Removal of breast skin, as performed by some surgeons during quadrantectomy for early forms of breast cancer, will lead to constriction of the breast contour in that area. Modeling to predict the impact of segmental skin resection in such cases would also be desirable.

The ability to predict breast contour and cosmesis after lumpectomy would be highly informative in patient decision-making. Preservation of the breast in patients with breast cancer has been shown to enhance the patient's sense of wholeness and recovery. While up to 85% of patients rate their own breast cosmesis as good or excellent, objective grading scales have demonstrated that results are more variable. The factors to consider in cosmesis are symmetry in size and shape to the opposite breast, texture of the breast, scar quality, and position of the nipple on the breast mound. If one could predict the cosmetic result that surgical lumpectomy would create, one could advise the patient in advance that a future oncoplastic procedure to achieve optimal cosmesis would be required. Oncoplastic soft tissue procedures are well-tolerated by patients and with advance knowledge that such refinements will be expected, patients can better prepare for the course of their surgeries with a positive perspective [14–16].

The Impact of Radiation Therapy on Breast Cosmesis

Radiotherapy is an essential component of therapy for patients undergoing BCT for cancer. Radiotherapy may be delivered via several techniques including external beam, intraoperative high dose, and catheter-based accelerated methods [17–21]. The tissue effects of radiation are evident in the skin and the breast parenchyma. An early inflammatory response characterized by lymphatic channel disruption and edema evolves with time to fibrosis and scarring, which includes firmness and

contraction in some patients. Factors predictive of significant radiation fibrosis are not defined. Predictive models of breast contour after BCT need to include the impact of radiation on the breast healing process, a process that continues to evolve over a several year period [22].

Prediction of Pathologically Negative Surgical Margins After Lumpectomy

A fundamental requirement of BCT is that the breast tumor is fully excised at surgery with microscopically negative margins. Given the pattern by which breast cancers grow, with speculated microscopic extension of tumor around a stellate core, the final assessment of microscopically negative margins cannot be determined until thorough pathological tissue analysis is completed after surgery, a process that can take several days to complete. The use of intraoperative frozen section analysis is prohibitive in all but unusual circumstances, as the time required, unjustifiably extends the patient's operation. As a result, microscopically positive margins – microscopic tumor present at the edge of the lumpectomy specimen – occurs in up to 20% of patients undergoing lumpectomy. A means to better predict the presence of microscopic disease is needed [23–25].

Two methods in computational surgery may apply. Modeling of tumor growth [26, 27] may allow one to define predictive patterns of tumor growth and extension within the matrix and anatomic structures of the breast as a function of tumor cellular and molecular features [28]. Cellular characteristics define cell–cell adherence, cell–matrix interactions, and cellular mobility. These factors will also be subject to the local anatomical environment governed by mesenchymal cell type, tissue density, the local growth factor milieu, and other features. Modeling of tumor growth based on known tumor molecular fingerprinting could provide a more accurate target to guide lumpectomy borders. This may be particularly true for ductal carcinoma in situ tumors whose growth is bounded and follows ductal structures of the breast [29]. Prediction of extension based on proliferative rate, estimated tumor volume, and other features that are below the threshold of current imaging modalities may offer improved surgical planning for lumpectomy.

Computational methods may also be useful in defining the volume or density of cancer cells that must be present at a margin. A variety of technologies are currently in development to improve intraoperative detection of positive margins. These include optical techniques, such as Raman spectroscopy and near-infrared reflectance confocal microscopy with silica-gold nanoshells [30, 31]. Raman spectroscopy has defined a spectroscopic model of breast tissue including the Raman spectra of epithelial cell cytoplasm, the cell nucleus, fat and other components of breast tissue. Preliminary ex vivo analysis of breast tissues demonstrate that the spectra can differentiate between normal and ductal carcinoma in situ in pre-clinical testing. Similarly, HER2-neu antibody conjugated silica-gold nanospheres have been shown to be detectable in a qualitative and quantitative manner on the surface of HER2-neu

expressing tumors in lumpectomy specimens. With further development, each of these technologies may allow more accurate and rapid intraoperative assessment of margins to facilitate one step lumpectomy in patients undergoing BCT.

Summary

Further advances in our understanding of breast cancer tumor biology, breast imaging, and technologies, to more accurately plan and execute surgical removal of malignant breast tissue will lead to improved cosmetic and oncologic outcomes for patients with breast cancer. The application of computational tools and mathematical modeling, linking these biologic and clinical phenomenon, offers the opportunity to advance the surgical management of patients with this common and highly treatable cancer that affects women around the world.

References

1. Miyoshi Y, Murase K, Oh K (2008) Basal-like subtype and BRCA1 dysfunction in breast cancers. Int J Clin Oncol 13:395–400
2. Huang E, Cheng S, Dressman H, et al (2003) Gene expression predictors of breast cancer outcomes. Lancet 361:1590–1596
3. Querci della Riovere G, Benson J (2002) Ipsilateral local recurrence of breast cancer: determinant or indicator of poor prognosis? Lancet Oncol 3:183–187
4. Veronesi U, Cascinelli N, Mariani L, et al (2002) Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med 347:1227–1232
5. American Joint Commission on Cancer USA (2005) AJCC staging manual, 7th edn. Am Joint Commission on Cancer USA
6. Benson J, Jatoi I, Keisch M, Esteva F, Makris A, Jordan M (2009) Early breast cancer. Lancet 373:1463–1479
7. Early Breast Cancer Trialists Collaborative Group (1998) tamoxifen for early breast cancer: an overview of the randomized trials. Lancet 351:1451–1461
8. Fisher B, Anderson S, Bryant J, et al (2002) Twenty-year follow up of a randomized trial comparing total mastectomy, lumpectomy and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 347:1233–1241
9. Early Breast Cancer Trialists Collaborative Group (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15 year survival: an overview of the randomized trials. Lancet 366:2087–2106
10. Morrow M, Harris J (2007) Practice guidelines for breast conserving therapy in the management of invasive breast cancer. J Am Coll Surg 205:362–376
11. Hata T, Takahashi H, Watanabe K, et al (2004) Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mammography and Ultrasonography. J Am Coll Surg 198:190–197
12. Bajaj A, Kon P, Oberg K, Miles D (2004) Aesthetic outcomes in patients undergoing breast conservation therapy for the treatment of localized breast cancer. Plast Reconstr Surg 114(6):1442–1449
13. Munshi A, Kakkar S, Bhutani R, Jalali R, Budrukkar A, Dinshaw K (2009) Factors influencing cosmetic outcome in breast conservation. Clin Oncol 21:285–293

14. Cochrane R, Valasiadou P, Wilson A, et al (2003) Cosmesis and satisfaction after breast conserving surgery correlates with percentage of breast volume excised. Br J Surg 90:1505–1509
15. Taylor M, Perez C, Halverson K, et al (1995) Factors influencing cosmetic results after conservation therapy for breast cancer. Int J Radiat Oncol Biol Phys 31:753–764
16. Behranwala K, Dua R, Ross G, et al (2006) The influence of radiotherapy on capsule formation and aesthetic outcome after immediate breast reconstruction using biodimensional anatomical expander implants. J Plast Reconstr Aesthet Surg 59:1043–1051
17. Vicini F, Beitch P, Quiet C, et al (2008) Three-year analysis of treatment efficacy, cosmesis, and toxicity by the American Society of Breast Surgeons MammoSite Breast Brachytherapy Registry Trial in patients treated with accelerated partial breast irradiation (ABPI). Cancer 112:758–766
18. Keisch M, Vicini F, Kuske R, et al (2003) Initial clinical experience with the MammoSite breast brachytherapy applicator in women with early-stage breast cancer treated with breast-conserving therapy. Int J Radiat Oncol Biol Phys 55:289–293
19. Veronesi U, Orrechia R, Luini A, et al (2001) A preliminary report of intraoperative radiotherapy (IORT) in limited-stage breast cancers that are conservatively treated. Eur J Cancer 37:2178–2183
20. Vaidya J, Baum M, Tobias J, et al (2001) Targeted intra-operative radiotherapy (Targit): an innovative Methodist of treatment for early breast cancer. Ann Oncol 12:1075–1080
21. Kestin L, Sharpe M, Frazier R, et al (2000) Intensity modulation to improve dose uniformity with tangential breast radiotherapy: initial clinical experience. Int J Radiat Oncol Biol Phys 48:1559–1568
22. Budrukkar A, Sarin R, Shrivastava S, Deshpande D, Dinshaw K (2007) Cosmesis, late sequelae and local control after breast-conserving therapy: influence of type of tumour bed boost and adjuvant chemotherapy. Clin Oncol 19:596–603
23. Pleijhuis RG, Graafland M, de Vries J, Bart J, de Jong JS, van Dam GM (2009) Obtaining adequate surgical margins in breast-conserving therapy for patients with early-stage breast cancer: current modalities and future directions. Ann Surg Oncol 16:2717–30
24. Jacobs L (2008) Positive margins: the challenge continues for breast surgeons. Ann Surg Oncol 15:1271–1272
25. Schiller D, Le L, Cho B, Youngson B, McCready D (2007) Factors associated with negative margins of lumpectomy specimen: potential use in selecting patients for intraoperative radiotherapy. Ann Surg Oncol 15(3):833–842
26. Enderling H, Vaidya JS (2008) Mathematical modelling of breast carcinogenesis, treatment with surgery and radiotherapy, and local recurrence, selected topics in cancer modeling. Birkhauser Boston, Cambridge, MA, pp 1–25
27. Ribba R, Colin T, Schnell S (2006) A Multiscale mathematical model of cancer and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7
28. Marchionni I, Wilson R, Wolff A, et al (2008) Systematic review: gene expression profiling assays in early stage breast cancer. Ann Intern Med 148:358–369
29. Hwang E, Kinkel K, Esserman L, et al (2003) Magnetic resonance imaging in patients diagnosed with ductal carcinoma-in-situ: value in the diagnosis of residual disease, occult invasion, and multicentricity. Ann Surg Oncol 10:381–388
30. Bickford LR, Agollah G, Drezek R, Yu TK (2009 May 6) Silica-gold nanoshells as potential intraoperative molecular probes for HER2-overexpression in ex vivo breast tissue using near-infrared reflectance confocal microscopy. Breast Cancer Res Treat [Epub ahead of print])
31. Haka A, Volynskaya Z, Gardecki J, Nazemi J, Lyons J, Hicks D, Fitzmaurice M, Dasari R, Crowe J, Feld M (2006) *In vivo* margin assessment during partial mastectomy breast surgery using Raman spectroscopy. Cancer Res 66(6):3317–3322

Changing Paradigms in the Management of Peripheral Vascular Disease: The Need for Integration of Knowledge, Imaging, and Therapeutics

Mark G. Davies and Daynene Vykoukal

Clinical Problem

Peripheral arterial occlusive disease (PAD) due to atherosclerosis of the lower extremities affects 3–7% of the population and up to one in five patients over 75 years of age. It is associated with decreased measures of quality of life [1–3] and is an underrecognized marker for multisystem atherosclerotic vascular disease. The risk of disease increases two- to threefold for every 10-year increase in age after the age of 40 years [4–6], with males developing claudication about twice as commonly as females. Mortality in patients with PAD is up to four times that of the nonPAD age-adjusted population [7]. Most (55%) die from heart disease, 10% from a stroke, and 10% from abdominal vascular pathology [8–12]. Less than 20% of PAD sufferers will die from a nonvascular cause. PAD is easily measured by the ratio of the systolic blood pressure in the upper arm and the systolic blood pressure at the ankle – the ankle brachial Index (ABI). The normal ratio is 1.0. The strength of association is so strong that even an asymptomatic patient with a slightly reduced ABI of 0.9 has a twofold relative risk of a coronary event [13]. Anatomic distribution of PAD is important. Patients with PAD can have disease in the aortic, iliac, femoral, and tibial vessels of the lower extremity. Patients with isolated aorto-iliac vessel disease tend to be younger and have a lower likelihood of pre-existing coronary heart disease. Those with femoral vessel disease, tibial vessel disease, or disease in all three vessels tend to have the lowest ABI and the highest likelihood of coronary heart disease [7, 14–19]. There is at present no current national US database on vascular interventions to allow realtime analysis of the trends in therapy and outcomes.

M.G. Davies (✉)
Methodist DeBakey Heart and Vascular Center, Department of Cardiovascular Surgery, The Methodist Hospital, Houston, TX, USA
e-mail: MDavies@tmhs.org

D. Vykoukal
The Methodist Hospital Research Institute, Houston, TX, USA
e-mail: DVykoukal@tmhs.org

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_2,

Anatomy

The superficial femoral and tibial arteries are muscular arteries. The superficial femoral artery is unique in that it is the longest artery in the body, and courses through the thigh in the muscular adductor canal, exiting at a fixed point. The tibials course through muscular skeletal compartments and are end arteries to the foot. The geometry and the elasticity of the SFA and tibials are significantly influenced by its proximity to musculature and its continuous mobility. They are further influenced by their location between two joints. Thus, they undergo additional mechanical forces not seen in other arteries. Both have unique elastic wall recoil properties that affect their conformability and resilience [20–22]. The flow in the SFA and tibial arteries also differs from that in many vessels that are treated with angioplasty, as they have high resistance characteristics and disturbed flow [23–25]. Nonlaminar flow results in increased predisposition to the development of atherosclerosis and intimal hyperplasia [26, 27]. One can model the wall and movement of the SFA, but little data has been published in this field. The ability to model such conditions would allow preoperative planning and intervention.

Pathophysiology and Classification of Disease

While atherosclerosis is a systemic disease affecting all arteries within the human body, the lower extremity arteries, because of their anatomy and physiology, are very susceptible to the development of atherosclerosis, and can demonstrate all aspects of atherosclerotic plaque development. Atherosclerosis may be considered as a space-occupying lesion within the arterial lumen that has significant influence on the structure of the vessel wall and is a dynamic biological structure undergoing inflammation, expansion, contraction, and remodeling, depending on both local and systemic conditions [28, 29]. Various types of lesions have been described in the fatty streak (Types I, II and III lesions), fibrofatty lesion (Types III, IV and Va,), and the fibrous plaque (Types Vb, Vc and VI) [28, 29]. The SFA and tibial vessels demonstrate all these lesions, but by the time the lesion is symptomatic (i.e., manifested by claudication), the fibrous plaque or advanced complicated lesion predominates, commonly with associated in situ thrombosis or occlusion [30]. The dynamic nature of the plaques has increased the interest in medical therapies to induce stabilization and/or regression in plaque morphology. Plaque growth, whether slow or abrupt, is usually associated with surface erosion, internal hemorrhage, luminal thrombosis, a combination of these processes, or gradual vessel occlusion [28, 29]. The most hazardous component of the atherosclerotic lesion is the plaque margin, where the fibrous cap of the lesion is thinned and eroded by macrophages and the metalloproteinases they secrete [31]. Sudden deaths from myocardial infarcts or acutely ischemic limbs are due to these ruptures or fissures in the margins of the fibrous cap [32]. Numerous clinico-pathologic investigations have demonstrated that surface weakness is the most common feature of

an unstable plaque [28, 29, 33]. Microscopically, the observed sites of injury span a broad morphologic range from minimal surface erosions to lacerations that extend deep within the plaque. The result of these injuries is exposure of the luminal blood to a thrombogenic surface, thereby setting the stage for acute thrombosis. Plaque rupture can also lead to hemorrhage within the atheroma [34]. Although the region of hemorrhage consists primarily of red blood cells, the surfaces along the ruptured tract are often lined by aggregates of platelets. Plaque hemorrhage can also develop by an entirely different mechanism. Within the core of a soft atheroma, primary disruption of capillary channels derived from the vasa vasorum may occur and lead to rapid plaque enlargement. Small hemorrhages are frequently observed in nonruptured plaques [34]. Mechanical stress points for eccentric lesions have been identified at the junction of the plaque with the more normal appearing arterial wall and at the center of the plaque [32–34]. In addition to the atherosclerotic lesion characterizations, there is an anatomical classification of disease based on angiography or alternative imaging modalities (duplex ultrasound, computed tomography angiography (CTA) or magnetic resonance angiography (MRA)). The TransAtlantic interSocietal Commission (TASC) has stratified femoropopliteal disease into four categories, A, B, C and D [35, 36]. These categories are used as an aid to guide intervention and for clinical reporting purposes. Integration of a bioinformatics approach to the analysis of plaque characteristics and the biomechanics of the SFA has not been achieved.

Pathology of Angioplasty and Stenting

Angioplasty is a controlled injury to the vessel wall. In the immediate aftermath of angioplasty, programmed cell death or apoptosis can be identified at 1–2 h and appears to disappear by 4 h [37]. No apoptosis can be identified in the wall after injury at 3 days, but by day 7, 50% of the cells again show signs of apoptosis and by day 14, the number of apoptotic cells is again markedly decreased. Smooth muscle cell proliferation within the media, which is normally less than 1%, increases to over 20% within 48 h after angioplasty [38–40]. The fraction of cells proliferating reaches a maximum between 3 and 7 days and occurs as a synchronous wave of entry into the S phase of the medial smooth muscle [41, 42]. Four weeks after injury, the medial proliferative response returns to baseline levels. Intracoronary radiation after angioplasty inhibits this first wave of cell proliferation and prevents adventitial proliferation [43]. By day 8 after the injury, smooth muscle cells are observed on the luminal side of the internal elastic lamina and appear to have migrated to the luminal surface through fenestrations in the internal elastic lamina. The number of smooth muscle cells in the intima increases to a maximum at 2 weeks after injury and about 30% of medial smooth muscle cells may migrate from the media to the intima. This migration of cells requires the proteolytic degradation of the cage of matrix surrounding each cell and the synthesis of new matrix molecules. Smooth muscle cell migration is unaffected by irradiation and anti-mitotic drugs

[44–47]. Once within the intima, approximately 50% of the smooth muscle cells proliferate (a second phase of mitosis). In the intima, a second phase of cellular proliferation is first noted at day 7 and reaches a maximum at 14 days before returning to baseline by 28 days [42]. However, it may continue for up to 12 weeks in those areas where re-endothelialization takes longer to complete. This second phase of smooth muscle cell replication in the intima appears to be mediated by autocrine and paracrine factors and remains poorly understood. It also appears that the thickness of the intimal hyperplasia peaks within 1 month and its rapid development is due to both cellular elements and the production of proteoglycans. Associated with the changes in the intima and media, there are substantial changes in the adventitia, as evidenced by an increased cell proliferation and growth factor synthesis in the adventitia relative to the media after angioplasty. In the adventitia, there is a marked infiltration of cells termed as "myofibroblasts" by day 2, which by day 14 can represent up to 50% of cells within the intima [48, 49]. The presence of myofibroblasts is common in wound healing and leads to contraction of the wound. A similar phenomenon may occur in the healing vessel. Injured vessels may undergo chronic elastic recoil or negative remodeling, which results in loss of luminal dimensions without a further increase in neointimal area. The degree of intimal hyperplasia that develops in a vessel is dependent on the length and depth of the injury [50]. The length of the injury influences the duration of the re-endothelialization process. Re-endothelialization occurs from the margins of the denuded area and possibly from the endothelial cells of the vasa vasorum. The longer there is an incomplete endothelial cell covering, the greater time the smooth muscle cells are without the modulating influence of the endothelial cells, and the longer the replication phases of the smooth muscle cells will be [39, 51]. After deep vessel wall injury, luminal narrowing may be less dependent on intimal hyperplasia formation and more dependent on vessel wall remodeling [52]. Medial damage is accompanied by a massive adventitial cell proliferation [53], which in time provides cells capable of contraction and negative remodeling. If angioplasty fails to achieve adequate luminal increase, causes vessel dissection or results in abrupt occlusion, intravascular stents are placed. At present, there is no systems biology approach to angioplasty, complications of angioplasty, and intravascular stenting (vide infra).

Intravascular Stents

The biology of in-stent restenosis is different than that seen after balloon angioplasty [54]. A stent is generally used if the result of balloon angioplasty is technically unsatisfactory or if there is arterial occlusion, immediate elastic recoil, flow-limiting dissection or restenosis. The response of a vessel to a stent is dependent on the stent design, length, composition, delivery system, and deployment technique [55]. In-stent restenosis is classified on the basis of length of restenosis in relation to stented length. Four categories of in-stent restenosis have been defined (1) focal ($\leq$10 mm length), (2) diffuse (>10 mm length), (3) proliferative (>10 mm length

and extending outside the stent), and (4) occlusion [56]. After balloon angioplasty, there is thrombus formation, intimal hyperplasia development, elastic recoil, and negative remodeling. In contrast, after stent placement, elastic recoil, and negative remodeling are eliminated [57] and thrombus formation followed by intimal hyperplasia development are the main contributors to in-stent restenosis [58, 59]. Stent placement in a vessel results in both a generalized injury to the length of the vessel exposed as well as the more focal injuries at the areas of stent placement. Intravascular ultrasound has demonstrated that stents do not always completely oppose the vessel wall along its entire length, thus resulting in uneven injury along its length [57]. After stent placement, the surface of the metal implanted into the vessel is covered within 5 s by a strongly adherent monolayer of protein. After 1 min, the surface is covered by fine layers of proteins, predominantly fibrinogen [60]. The holes between the stent wires are filled with thrombus and the adherence of platelets and leukocytes is enhanced by disturbance of the electrostatic equilibrium [61, 62]. The basic mechanisms of smooth muscle cell proliferation and migration after stent placement are the same as those after balloon injury [63]. The intimal hyperplasic process in a stent is more prolonged and robust than in a balloon injured artery and is proportional to the depth of injury the recipient vessel sustains [64] and the inflammatory response induced [65]. It can often be much more significant at the ends than in the body of the stent. In addition, the adventitial response is prolonged, with adventitial giant cell body formation being noted. Stents prevent chronic elastic recoil and cause progressive atrophy of the media [66]. The presence of a stent changes the dynamics of the vessel wall both at the site of implantation and distally by changing the flexibility of the entire vessel. The stresses on the stent have led to fractures of the stent and these fractures are increased by increased patient mobility. A combination of systems biology and CAD may allow a better definition of the causes and possible remedies for these stent related problems.

Diagnostic Imaging

Imaging of the arterial system can be achieved through the use of Duplex ultrasound (Doppler and B-mode ultrasound combined), computed tomography arteriography, magnetic resonance angiography and conventional arteriography with the use of intravascular ultrasound modalities. Duplex allows real-time imaging, integrating blood flow, and two-dimensional imaging and can quantify flow, degrees of obstruction, and plaque characteristics. When duplex ultrasound (DUS)-derived gray-scale median (GSM) is used to interrogate the distal portion of an occluded native femoral-popliteal arterial segment, it can predict the success of lumen re-entry for subintimal angioplasty. When the GSM was less than 25, the absolute reduction in plaque thickness on day 1 post-percutaneous transluminal angioplasty (PTA) was 3.3 ± 1.8 mm, in contrast to 1.8 ± 1.6 mm when GSM was more than 25 ($P < 0.03$) [67]. The restenosis rate (peak systolic velocity ratio (PSVR) more than 2) was 41% at 6 months and remained unchanged at 1 year. When the GSM was less than 25, restenosis occurred in 11% of lesions, in comparison with 78% when the GSM was

more than 25 ($P < 0.001$). Failure to recannulize occurred in 90% of 19 cases with GSM > 35, in 71% of 24 cases with GSM > 20, and in 50% of 34 cases with GSM > 25. Plaque echogenicity represented by DUS-derived GSM can be used to predict the success of primary subintimal femoral-popliteal angioplasties [68].

Overall, the sensitivity of Computed tomography arteriography (CTA) for detecting more than 50% stenosis or occlusion was 95% (95% confidence interval [CI], 92–97%) and specificity was 96% (95% CI, 93–97%). CTA correctly identified occlusions in 94% of segments, the presence of more than 50% stenosis in 87% of segments, and absence of significant stenosis in 96% of segments. Overstaging occurred in 8% of segments and understaging in 15% [69]. The diagnostic performance of multidetector CT angiography in the infrapopliteal tract is lower than but not significantly different from that in the aorto-iliac and femoropopliteal tracts. Regression analysis showed that diagnostic performance was not significantly influenced by differences in study characteristics [70]. When evaluated by TASC classification, time-resolved imaging of contrast kinetics (TRICKS) MRA correlated with digital subtraction angiography (DSA) in 83% of the popliteal and in 88% of the infrapopliteal segments. MRA correctly identified significant disease of the popliteal artery with a sensitivity of 94% and a specificity of 92%, and of the tibial arteries with a sensitivity of 100% and specificity of 84% [71]. Intra-arterial (IA) contrast-enhanced 3D-gradient-echo-MRA (CE MRA) on an open-bore MR-scanner with the detection of significant stenoses and occlusions, the overall sensitivity, specificity, positive predictive value, negative predictive value, and accuracy of IA-MRA were 96, 83, 88, 94, and 90%, respectively [72]. A recent meta-analysis sought to determine the diagnostic accuracy and cost-effectiveness of duplex ultrasound (DUS), magnetic resonance angiography (MRA), and computed tomography angiography (CTA), as alternatives to contrast angiography (CA), for the assessment of lower limb peripheral arterial. The results of the review suggested that CE MRA has a better overall diagnostic accuracy than CTA or DUS, and that CE MRA is generally preferred by patients over CA. Where available, CE MRA may be a viable alternative to CA. There was insufficient evidence to evaluate the usefulness of CTA for the assessment of PAD, particularly newer techniques. The results of the economic modeling suggest that for PAD patients for whom the whole leg is evaluated by a preoperative diagnostic test, DUS dominates the other alternatives by presenting higher effectiveness at a lower cost per quality-adjusted life year (QALY). However, when the analysis of stenosis is limited to a section of the leg, either above the knee or below the knee, 2D time-of-flight (TOF) MRA appears to be the most cost-effective preoperative diagnostic strategy. Further research is needed into a number of areas including the relative clinical effectiveness of the available imaging tests, in terms of surgical planning and postoperative outcome [73].

Angiography provides information on luminal characteristics of peripheral arteries but severely underestimates the extent of atherosclerosis in patients with PAD even in “normal appearing” vessels [74]. When plaque composition in the popliteal and tibial vessels of patients with severe peripheral arterial disease was examined by intravascular ultrasound (IVUS) virtual histology, tibial vessels had more dense plaque calcium than popliteal arteries. Consequently, distal vessels had less fibro-fatty and fibrous plaque than popliteal arteries. Necrotic core plaque

composition was found to be similar when comparing tibial vs. popliteal arteries. Clinical factors including diabetes, hyperlipidemia, and chronic renal insufficiency were not associated with plaque composition differences using a univariate analysis [75].

If IVUS is used immediately after PTA, the extent of dissection, free lumen area, and diameter are predictive factors of patency [76]. Factors on IVUS that favor improved patency are the absence of calcification, dissection or plaque rupture and a residual stenosis of less than 30% [77]. IVUS predictor of failure at one and 6 months is initial residual stenosis after PTA [78]. Similarly, significant residual stenosis (>30%) on duplex ultrasound 1 day after PTA correlates with failure within 1 year; unfortunately, the converse is not true, as a normal duplex at +1 day cannot predict failure within 1 year [79]. Plaque area increase and vascular remodeling contribute to lumen area change after PTA of the femoropopliteal artery on intravascular ultrasound study [80]. Capek [81] and Hunink [82] have shown that repeat PTA has similar patency as primary PTA; in contrast, several reports have shown that repeat SFA angioplasty has a low patency (<20% at 3 years) and have suggested that bypass may be more beneficial [83–86]. Failure of PTA is not associated with worsening of the patient's clinical condition and 19% will be recannulized at a second sitting [87, 88].

The graduation of femoropopliteal stenoses by either digital subtraction angiography (DSA) or duplex sonography remains challenging, particularly after percutaneous transluminal angioplasty (PTA). The relationship between DSA, IVUS and duplex before and after femoropopliteal PTAs has been examined. Over the whole range of stenoses, peak systolic velocity (PSV) and peak velocity ratio (PVR) correlated better with DSA-stenosis ($R(2) = 0.72$ and 0.74, respectively; $P < 0.01$) than with IVUS-stenosis ($R(2) = 0.58$ and 0.50; $P < 0.01$). Within the subgroup of preinterventional (51–99%) stenoses, PVR was significantly correlated only with DSA-stenosis ($R(2) = 0.60$; $P < 0.01$). Severe dissection after PTA was associated with a disproportionate rise in PSV and large discrepancies between IVUS and DSA. Of note, it has been reported that intra-stenotic flow acceleration assessed by duplex sonography correlates better with DSA- than with IVUS-stenosis. The concordance between duplex sonography, DSA, and IVUS was particularly weak in postinterventional measurements, casting some doubt on the reliability of these methods for the assessment of residual stenosis after femoropopliteal PTA [89]. Integration of various image modalities with simulation and modeling which is then individualized to the patient is a growing area of interest. Integrated registration of the cumulative imaging a patient receives is necessary to enhance medical decision making and intervention.

Endoluminal Procedures

Since the introduction of percutaneous transluminal angioplasty in 1964 by Dotter and Judkins [90] and its refinements by Gruntzig [91], PTA has grown into a viable alternative to surgical intervention in most arterial beds. The diseased SFA may be

approached with a repertoire of techniques. The conventional approach is to access the SFA in an antegrade manner and negotiate a guide wire through the target lesion prior to performing PTA. A similar approach can be achieved in short segment occlusions. An antegrade transfemoral approach has initial technical success rates of 90–95% for stenotic lesions and of 80–95% for complete occlusions. Use of the retrograde transpopliteal approach can increase the original technical success rate by an additional 6% [92]. However, longer occlusions require additional techniques. Percutaneous intentional extraluminal recanalization (PIER) [93] or sub-intimal angioplasty or Bolia angioplasty is a technique used to recannulize chronic occlusions [94]. It intentionally achieves a dissection plane in the vessel wall to circumvent a total occlusion. Gaining luminal access into the distal target vessel remains the Achilles' heel of this approach. Results from subintimal angioplasty of superficial femoral artery occlusions were superior to the results of PTA, but calcification is associated with subintimal angioplasty failure [95]. Recently, the use of intravascular ultrasound to identify and guide a needle from the false dissected lumen into the true lumen has been shown to be effective. PIER through the retrograde popliteal approach can be achieved in up to 80% cases [96].

Outcomes

The national reporting standards for vascular procedures define three categories of clinical success: anatomic, hemodynamic, and clinical. A composite analysis of all published trials of PTA with or without stenting in the femoropopliteal arteries shows patencies of 71, 59, and 53% at 1, 3, and 5 years, respectively. Technical success was 90% and the complication rate was 10%. Endoluminal therapy for SFA occlusive disease yields lower assisted patency rates and higher restenosis rates for those patients presenting with claudication who have insulin dependent diabetes. Among those patients presenting with tissue loss, limb salvage rates are lowered for all diabetics (both noninsulin-dependent and insulin-dependent diabetes mellitus (NIDDM and IDDM)) in spite of equivalent patency and restenosis rates [97]. Insulin levels and the C-peptide/insulin ratio were associated with restenosis after femoral PTA. In nondiabetic patients, insulin levels were significantly associated with restenosis, whereas the ratio of C-peptide to insulin showed no association with restenosis. In patients with type II diabetes, in contrast, the C-peptide/insulin ratio was associated with restenosis, whereas insulin levels showed no significant association with restenosis [98]. Multiple factors may adversely affect patency; these include presenting symptoms (claudication vs. critical ischemia), type of lesion (stenosis vs. occlusion), length of lesion <10 and >10 cm) and distal runoff [81, 82, 99, 100]. Additional significant predictive factors of a good outcome are as follows: female gender, nondiabetic, at least one patent artery below the knee, AHA classification <2, no stent, treatment of an occlusion, number of dilatations <3, treatment by statins for hypercholesterolemia [101]. The IL-6 promoter polymorphism (-174)G/C seems to influence the occurrence of restenosis after PTA. Homozygous carriers of the (-174)C allele have an increased rate

of intermediate-term restenosis [102]. Plasma levels of tissue factor, prothrombin fragment 1 + 2, D-dimer, P-selectin, C-reactive protein (CRP), and fibrinogen analyzed before and after angioplasty are not related to restenosis [103]. Plasma tissue factor is a predictor for restenosis after femoropopliteal angioplasty [104].

Subintimal angioplasty for the treatment of lower extremity chronic arterial occlusions is technically feasible, results in minimal morbidity, and provides satisfactory revascularization without surgical bypass. Secondary patency is comparable to that of autologous vein bypass and is achieved with a low rate of reintervention [105, 106]. However, percutaneous SFA intervention in preparation for a distal origin graft is a useful and effective strategy in select patients. The durability appears comparable with distal origin grafts performed in the absence of an SFA intervention. This strategy provides a good option in the setting of both atherosclerotic SFA disease and limited autogenous conduit. Simultaneous performance is not associated with increased morbidity and decreases overall hospital use [107, 108]. Multivariate analysis reveals critical limb ischemia to be the only predictor of reduced primary patency [106]. Most amputations occurred in patients in whom subintmal angioplasty had been unsuccessful and were associated with long ($>$ or $=$ 10 cm) occlusions.

Pharmacotherapy After PTA

The role of pharmacotherapy after PTA in the SFA has been driven by the data derived from the coronary artery interventional literature. Heiss et al. have shown a significant benefit of aspirin compared to placebo after PTA [109], but antiplatelet drugs have not shown a consistent benefit in preventing restenosis [109, 110]. Low dose aspirin appears as effective as high dose aspirin in the prevention of restenosis [111]. The coronary literature supports the use of the combination of aspirin and clopidogrel after endoluminal intervention. No data is readily available in the periphery; however, most centers add clopidogrel prior to and after intervention. The required duration of this additional therapy has not been tested. Adjunctive abciximab, a platelet receptor inhibitor, does not appear to demonstrate any identifiable effect [112]. Statins can decrease restenosis and the reintervention rate after percutaneous intervention, independent of their lipid-lowering effect and CRP level in this study. Multivariate analysis indicated that the prescription of statins, but not LDL-C level at follow-up and % reduction of LDL-C during the follow-up period, predict the restenosis prevention [113] One small study compared changes in IVUS morphology of the lumen, vessel, and plaque in statin treated ($n = 5$) and non-statin-treated patients ($n = 5$). At follow-up, both statin-treated and nonstatin-treated patients showed a similar increase in plaque volume at the nondilated segment (+4% and +2%, respectively). In statin-treated patients, the plaque volume increase was compensated by an increase in vessel volume (+2%), resulting in an increase in lumen volume (+1%). In non-statin-treated patients, on the other hand, the increase in plaque volume was associated with a decrease in vessel volume (−2%), resulting

in a decrease in lumen volume (−4%) [114,115]. Management of comorbidities and medication dosing reconciliation and monitoring are integral to effective electronic medical management.

Adjuvant Stenting

Four randomized studies comparing PTA alone vs. PTA plus stent placement in the SFA have all failed to demonstrate a benefit to stenting in terms of long-term patency and symptom relief [116–119]. There is little evidence to support the superiority of endovascular stents over PTA alone in the femoropopliteal arteries. Their use should be confined at present to flow limiting dissections or inadequate results from balloon angioplasty alone [117,120,121]. Cheng et al. have suggested that significant factors which impede stent patency were occlusions, stented segment length >10 cm, use of the procedure in claudicants and the use of Memotherm stents [122]. Liermann reported a primary patency rate of 82% after 18 months and noted a lower patency rate for stents placed in the distal portion of the SFA compared to the proximal portion [123]. Strecker et al. reported a 2-year patency rate of 73% for stent placement for stenosis and 32% in occlusions [124]. Use of wall stents for short lesions and stenoses reveals a 1–2-year primary patency of 67%, while for occlusions this dropped to 49% [125]. One report has suggested a high rate of early thrombosis and a restenosis rate of 77% when wallstents are placed in the SFA [126]. One report has shown that the 12-month restenosis of nitinol stents is 46%, which was not influenced by the cumulative length or the number of stents placed, but was worsened in the presence of diabetes [127]. Self-expanding stents produce acceptable outcomes for the treatment of SFA disease. Poorer patency rates are associated with TASC D lesions and poor initial runoff score; hypertension was associated with improved patency rates. Stent occlusion and in-stent stenosis were not entirely benign; one-third of patients had deterioration of their tibial artery runoff [128]. Re-occlusion rates of 32–43% in the first 12–18 months after implantation have been observed [129–132]. The Femoral Artery Stent Trial (FAST study) was designed to investigate the impact of nitinol stenting of superficial femoral artery lesions with a maximum length of 10 cm on restenosis and clinical outcomes at 1 year. There was no difference in binary restenosis at 1 year between the implantation of a single Luminex nitinol stent and stand-alone PTA [133]. The impact of nitinol stenting of superficial femoral artery (SFA) lesions with a maximum length of 10 cm (TASC-II A or B) on 1-year outcomes was compared to a historical study cohort from FAST was examined in the Femoral Artery Conformexx Trial (FACT study) [134]. Single de novo >70% SFA lesions <10 cm long treated with the self-expanding nitinol Conformexx stent had a primary endpoint of ultrasound-assessed binary restenosis which was lower vs. the historical balloon angioplasty. Reconstructive surgery in patients presenting with infrainguinal stent occlusion or restenosis appears to be associated with higher morbidity and major limb amputation rates [135]. Two recent meta-analyses suggest that stent placement in the femoropopliteal occlusive disease does not increase the patency rate when compared with angioplasty alone at 1 year [136,137].

Drug Eluting Stents

While several drug eluting stents have been utilized in the coronary circulation, only one, the sirolimus eluting SMART stent, has data available in the periphery. The results of the SIROCCO studies show that despite excellent 6-month patency observed in the bare nitinol stent group, comparative analysis reveals a reduction in neo-intimal hyperplasia in the sirolimus eluting stents [138–140]. It does not appear that these stents had a significant clinical effect. The final report stated that the cumulative in-stent restenosis rates according to duplex ultrasound were 22% at 24 months and that no significant difference could be found between the sirolimus-eluting and the bare SMART stents [141].

Covered Stents

In a bid to counteract in-stent restenosis, covered stents have been proposed [142]. There appears to be an initial gain in patency over the first 6–12 months [143, 144]. While this strategy did prevent in-stent restenosis, it did not prevent the development of intimal hyperplasia at the ends of the device [129]. The long term patency of the Viabahn stent-graft for long Transatlantic Inter-Society Consensus class C and D lesions was 79% at 4 years and was independent of lesion length and type but was dependent on device diameter. The primary vessel patency rate in devices of at least 7 mm ($n = 21$) was 82% at 4 years. No stent fractures were observed despite the use of multiple overlapping stent-grafts [145]. Severity of lesions, rather than preoperative symptoms or runoff, is mainly to be considered before using Hemobahn/Viabahn endoprosthesis in severe SFA occlusive lesions, as primary and secondary patencies decrease as TASC severity worsens [146]. Treatment of SFA occlusive lesions (excluding TASC D lesions) with the Hemobahn covered stent yielded good results for both claudicants with good outflow and patients with critical or acute ischemia with bad outflow if concomitant outflow-improving procedures were performed [147]. Results from a prospective multicenter registry evaluating the safety, effectiveness, and patency of the aSpire self-expanding polytetrafluoroethylene covered stent showed good mid-term results, but with a number of reinterventions necessary to obtain an optimal secondary patency. Risk of patency failure was related to critical limb ischemia as an indication for the procedure. Technologic and pharmacologic improvement and longer follow-up are needed to define the indications for the aSpire stent [148].

Atherectomy

Debulking of lesions may enhance patency. Atherectomy devices and endovascular endarterectomy have been used to achieve these aims. Directional atherectomy appears to have a 75% patency at 24 months but, there is a high restenosis rate [149].

Endovascular SFA endarterectomy appears to have a 65% technical success rate and a cumulative patency of 59% in patients, in whom technical success can be achieved [150]. Directional atherectomy has a patency of 57% at 2 years with patients with diabetes, complete luminal occlusion or critical ischemia having lower patency [151]. Studies have suggested that debulking segments prior to adjunctive balloon angioplasty may offer advantages in reducing acute complications and improving long term patency [152, 153]. In addition, treatment of in-stent restenosis may be helped by debulking prior to dilation [154]. Several mechanical atherectomy devices have been used in the treatment of peripheral vascular disease. The Simpson Atherocath is commonly employed to treat focal eccentric lesions in larger vessels and bypass graft anastomoses. It is limited by its high profile and side-cutting design. The SilverHawk catheter (FoxHollow Technologies, Inc.), which is derived from the Simpson catheter, has reduced these obstacles to catheter placement and use. Case series suggest that it has short-term benefits [155–157]. Rotational atherectomy (Rotablator) uses a very high speed rotating diamond coated burr to abrade lesions into microparticles which are dispersed down stream; it is very useful for heavily calcified lesions, but the consequences of distal embolization remains a significant concern. The TEC catheter uses aspiration to collect debris shaved by a rotating cutting head and appears more effective in treating soft or thrombus containing obstructions. In a retrospective analysis of 35 SFAs treated with the TEC device, the initial success rate was 100%, but the 5-year patency was 0% [158]. Both the pullback atherectomy catheter and the Redha-Cut catheter rely on first passing the obstruction with the entire device then removing the material as the catheter's cutting blades are pulled back across the lesion. In three prospective randomized trials, there was no difference between angioplasty and atherectomy for SFA disease [159–161]. Initial ABI, but not tibial runoff, or long occlusions, appeared to be the most predictive of patency after atherectomy. Lower extremity atherectomy procedures with the SilverHawk or Diamondback devices are a safe and effective means of improving symptoms. However, there is decreased durability, significant patency, and limb loss over time. Multivariable analysis demonstrated that tobacco use, renal disease, diabetes, and tissue loss are all predictors of patency loss, while only diabetes and tissue loss were associated with greater limb loss. There was no difference in patency rates irrespective of location, TASC classification, vessel treated (femoral vs. tibial), or degree of stenosis (occluded vs. stenotic) [162]. Several experiences suggest a very poor midterm patency of excisonal atherectomy, although a 74% limb salvage rate was maintained through secondary interventions [163].

Laser Assisted Angioplasty

Laser-assisted angioplasty increases luminal cross-sectional area by both atheroablation and vessel expansion without calcium ablation [164, 165]. Technical success with laser-assisted angioplasty is as high as 91% [166–168]. There appears to be better initial recanalization with laser-assisted angioplasty compared

to PTA, but no long-term differences between PTA and laser-assisted angioplasty have been appreciated [169, 170]. SFA lesions, including in-stent restenosis (ISR), can be treated in the majority of cases with directed laser atherectomy, significantly reducing plaque burden as measured by IVUS. Embolization was attributed to device-specific features of the prototype antegrade sheath design, which was discontinued [171]. Enclosed thrombolysis associated with PTA has no advantage over PTA alone [172].

Brachytherapy

Delivery of radiation to the injured field in a vessel has had beneficial effects. Endovascular radiotherapy can prevent intimal hyperplasia after stent implantation in femoropopliteal arteries [123] but partially prevents healing of disrupted vessel surfaces [173]. Gamma radiation induces positive vascular remodeling after balloon angioplasty, as shown in a prospective randomized IVUS scan study [174]. When doses of 12–14 Gy (gamma radiation) are applied endoluminally to the femoropopliteal segment after PTA, there is a twofold increase in 12-month patency compared to untreated controls [175–177]. Similar results were noted in the PARIS trial [178]. The Vienna-2-trial showed a significant reduction of the restenosis rate compared to controls. Subgroup analysis showed a significant decrease in restenosis rate in the subgroups with restenosis after former PTA occlusion and PTA length of greater than 10 cm. Significant reduction was not achieved in diabetes patients [179]. Endovascular brachytherapy (14 Gy) applied by an ^{192}Ir source to the vessel wall resulted in an absolute risk reduction of significant restenosis of 9%, yet in patients with totally occlusive disease this reduction was 32% [180]. Results suggest that intravascular brachytherapy is effective at improving the patency of femoropopliteal arteries undergoing PTA in the short term, particularly in nondiabetics with long occlusions (>10 cm) [181]. Restenosis and inward remodeling after PTA are delayed by intravascular brachytherapy. At 24 months, patients treated with brachytherapy have a larger lumen than those treated with PTA alone. The decrease in luminal and total vessel area between 3 and 24 months after intravascular brachytherapy indicates that the restenotic and remodeling process is not abolished but merely delayed with this therapy [182].

Cryotherapy

Cryoplasty is a novel therapy that combines conventional balloon angioplasty with application of cryotherapy (Fig. 1). Experimental data suggests that cryotherapy induces early arterial wall cell loss through apoptosis but does not alter intimal hyperplasia development or eventual lumen area compared to conventional balloon angioplasty [183]. The single system on the market at present uses a double balloon system that exerts 8 atm. on the lesion and is accompanied by cold-induced

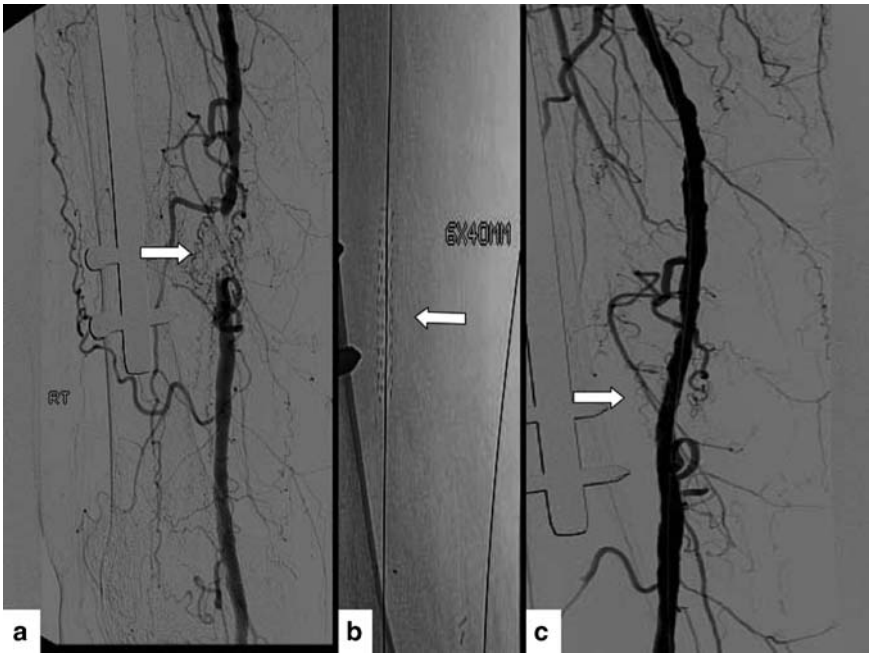


Fig. 1 Lower extremity angiogram showing in (**a**) a occlusion of the superficial femoral artery (*arrow*), (**b**) inflation of a Polarcath balloon angioplasty catheter (*arrow*), and (**c**) completion angiogram showing restoration of flow (*arrow*)

injury to the wall. The hope is that cryotherapy will allow more accurate angioplasty and induce apoptosis in the wall, thus reducing dissections and vessel response to injury. The final results of a cryovascular safety registry of 102 patients have shown a primary patency of 82% at 10 months [184, 185]. Longer-term data are required, although case series suggest that improved patency can be achieved for up to 18 months [186]. However, the mode of failure is different, with stenosis, more common than occlusion [187]. Associated costs are greater than those for conventional balloon angioplasty but lower than those for stent placement [187, 188]. The benefit of cryoplasty over conventional angioplasty has not been established, as no randomized controlled trials exist to properly evaluate this method. The technical success and primary patency rates seen in the prospective series are encouraging and may suggest a future role for cryoplasty in the treatment of PAD, but this data cannot be reliably interpreted due to the nature of the studies [189].

Cutting Balloon PTA

In a study of predominantly TASC A and B lesions, the overall technical success rate of a cutting balloon PTA (CB-PTA) was 96.3% and the complication rate was 8.9%. The 1- and 2-year results for femoropopliteal and infrapopliteal lesions in patients

with critical limb ischemia (CLI) were as follows: primary patency 64.4 and 51.9%, respectively; limb salvage 84.2 and 76.9%; survival 92.6 and 88.5%. More distal lesions and a high TASC classification were significant independent risk factors for outcome. Treatment of multiple segment lesions was an independent predictor of a favorable outcome ($P = 0.04$) [190]. Cotroneo and colleagues compared midterm results of cutting balloon angioplasty (CBA) to conventional percutaneous transluminal angioplasty (PTA) for the treatment of short femoropopliteal arterial stenosis. In the PTA group, primary and secondary patency rates, respectively, were 91.0 and 95.5% at 6 months, 83.1 and 92.4% at 12 months, and 66.6 and 76.5% at 2 years. In the CBA patients, the primary and secondary patency rates, respectively, were 93.2 and 95.9% at 6 months, 90.4% ($P < 0.001$ vs. PTA at same interval) and 94.5% at 12 months, and 79.7% ($P < 0.001$) and 85.6% ($P < 0.001$) at 2 years [191].

Complications of Intervention

In situ thrombosis appeared to occur more frequently in females, reflecting a smaller vessel size, current smokers with critical limb ischemia, and more complex procedures with longer lesions associated with poorer runoff that needed more than a simple angioplasty. While pharmaco-mechanical therapy did reopen the majority of these lesions, they did fail more frequently in follow up. In early reports, risk factors for acute thrombosis in the coronary circulation included stenting for threatened or abrupt closure, smaller vessels and longer lesions, and more recent studies have shown the risk factors to have shifted to multiple stent use, residual dissection, and smaller final lumen [192] (Fig. 2).

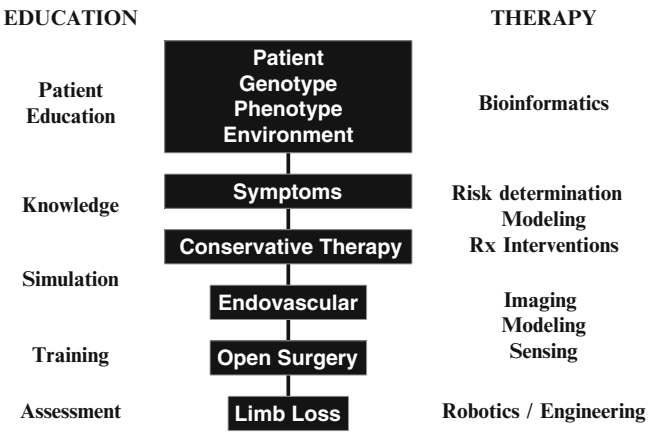


Fig. 2 Flow chart of patient care with computational interactions in the areas of education and therapy

Distal embolization has been identified during wire crossing, angioplasty, stent deployment, and atherectomy [193]. The frequency of embolization is greatest during stent deployment. The occurrences of embolic signals were particularly frequent in the 2 h after angioplasty [194]. In a single-center prospective registry (PROTECT), macro-embolization occurred in 55% of patients. Clinically significant (≥2 mm in diameter) macrodebris was found in 45% of patients [195]. In the PROTECT study and in a second smaller case series study, superficial femoral and popliteal artery atherectomy was associated with debris in the filter in all cases [195, 196]. When the particles captured in embolic filters were examined, the particles were found to consist primarily of platelets and fibrin conglomerates, trapped erythrocytes, inflammatory cells, and extracellular matrix. Increased lesion length, increased reference vessel diameter, acute thromboses, and total occlusions have been positively correlated with higher amounts of captured particles ($P < 0.05$). However, by multivariate analysis, it appears that declotting procedures were the only independent predictor of increased embolic burden in the study [197].

The criteria for embolic protection device placement in the PROTECT registry, and by implication, the criteria for distal embolization were: (1) moderate or severe calcification of any length, (2) total occlusions of any length, (3) a filling defect, (4) irregular (ulcerated) lesions at least 30 mm in length, (5) smooth, non-ulcerated lesions at least 50 mm in length. Distal embolization is associated with a greater likelihood of a major amputation, but this finding can be confounded by the fact that the patients initially presented with critical ischemia, complex lesions, and poor distal runoff.

Exercise vs. Angioplasty

Various trials have compared SFA PTA to exercise therapy. They suggest that although PTA may give a short 3- to 6-month improvement in symptoms, exercise therapy has better results at 12 months. However, at the 5-year follow up, there appears to be no differences between the groups. The loss of the initial advantage of PTA was attributable to restenosis of the dilated segment, whereas those treated conservatively had a steadily increased benefit due to exercise. The later loss of the advantage of exercise was probably due to a lack of adherence to the exercise program. The best effect of exercise was found in patients with disease confined to the superficial femoral artery, in contrast to PTA, which had the best effect in iliac lesions. One study has compared PTA to exercise to no treatment and has suggested that PTA moderately but significantly improves quality of life at 1 year. All these studies lacked sufficient power and the seminal group of PTA with supervised exercise is missing in all studies. Intervention for SFA occlusive disease does not provide superior long-term benefits compared with conservative medical therapy, with adverse events were observed more frequently in the intervention group than the control group (69.1 vs. 46.2%, $P < 0.05$). This was mainly due to

a higher frequency of re-hospitalization in the intervention group than in controls (52.7 vs. 15.4%, $P < 0.001$) [198]. In a recent randomized study, there was no significant difference in effectiveness between endovascular revascularization compared to supervised hospital-based exercise during 12-month follow-up, any gains with endovascular revascularization found were non-significant, and endovascular revascularization costs more than the generally accepted threshold willingness-to-pay value, which favors exercise [199].

Surgery vs. Angioplasty

When examining the benefits of exercise in the claudicant population, a further question arises: "is there a functional benefit of surgery in the intermittent claudication population?" One study has compared peripheral bypass surgery, surgery followed by 6 months of supervised exercise training with dynamic leg exercises, and 6 months of supervised training alone in 75 patients with claudication. At 13 months of randomization, walking ability was improved in all three groups. The most effective treatment for improving functional status was exercise training plus surgery. The median changes in maximal walking distance were 173% for the surgical group, 263% for the group that received both supervised exercise and bypass, and 151% for the exercise group alone. The surgically treated patients increased their walking distance more than patients who received only exercise training; they also had a greater rate of complications than the exercise groups. Percutaneous balloon angioplasty/stenting (PTA/S) for TASC-II C lesions has a superior midterm patency to that of above the knee femoropopliteal bypass (AK-FPB) using polytetrafluoroethylene (PTFE), and AK-FPB with PTFE has better primary and assisted-primary patency than PTA/S-D [200].

Economics

In Sweden, the cost effectiveness ratio (cost per month of patency) for PTA and local open thrombo-endarterectomy appears equivalent. For a decision and cost effectiveness analysis of revascularization procedures for femoropopliteal disease (4,800 PTAs and 4,511 bypasses), six treatment strategies were analyzed: (1) no treatment, (2) initial PTA with no further revascularization, (3) initial PTA with subsequent PTA, (4) initial PTA with subsequent bypass surgery, (5) bypass surgery followed by no therapy and (6) bypass surgery followed by graft revision. The results showed that for a 65-year-old man with disabling claudication and a femoropopliteal stenosis or occlusion, an initial PTA strategy increased QALY by 2–13 months and resulted in decreased lifetime expenditures as compared with bypass surgery. Sensitivity analysis showed that when the 5-year patency of PTA exceeds

30%, PTA is the preferred initial invasive strategy in patients with disabling claudication and femoropopliteal stenosis or occlusion. Endovascular revascularization of SFA disease improves quality of life (QoL), and restenosis negatively affects QoL outcomes. After stent implantation, whether primary or secondary, QoL was significantly ameliorated compared to balloon angioplasty alone. However, it remains to be proven in larger cohorts whether primary stenting yields a QoL benefit compared to balloon angioplasty with optional secondary stenting [201].

Conclusion

The superficial femoral artery (SFA) is the longest artery in the body and experiences a unique array of biomechanical forces during daily life. While the modeling and imaging of these vessels has been restricted to healthy volunteers, there is only limited data on the biological and mechanical features of these vessels as they develop atherosclerosis, as they stenose and produce symptoms, and as they respond to intervention (drug therapy, balloon angioplasty or stent placement). There are several highly relevant potential areas of research for computational sciences including bioinformatics, modeling vessel biomechanics, enhancing image interpretation, visualizing the impact of intervention on the vessel, and predicting the patterns of restenosis, dissection and occlusion after intervention (Fig. 3). Successful integration of clinical information with imaging, modeling, and simulations of the SFA is crucial to further therapeutic advances in the treatment of lower extremity disease (Fig. 4).

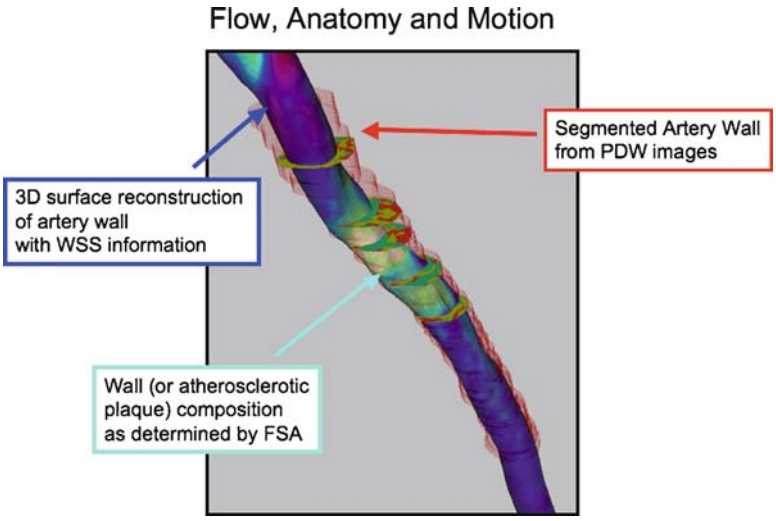


Fig. 3 Composite computational analysis of the SFA illustrating the potential of the integration of anatomy, composition and hemodynamics

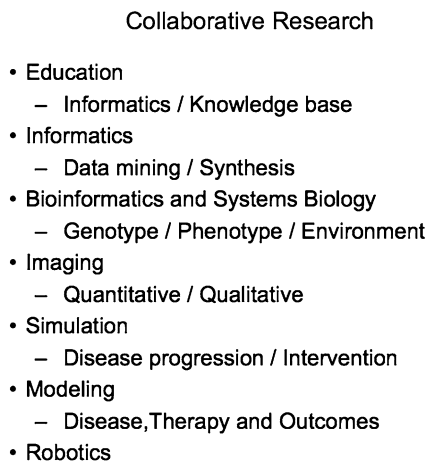


Fig. 4 Areas of collaborative research between clinical vascular surgery and computational sciences to achieve computational surgery

References

1. Jeremy M, Perkins T, Collin J (1996) Do exercise programmes improve claudication? In: Greenhalgh RM (ed) Trials and tribulations of vascular surgery. WB Saunders, Philadephia, PA, pp 259–267
2. Pell JP (1995) Impact of intermittent claudication on quality of life. Eur J Vasc Endovasc 9:469–472
3. Breek JC, Hamming JF, DeVries J, van Berge Henegouwen DP, VanHeck GL (2002) The impact of walking impairment, cardiovascular risk factors, and co-morbidity on quality of life in patients with intermittent claudication. J Vasc Surg 36(1):94–99
4. Criqui MH, Fronek A, Barrett-Conner E et al (1985) The prevalence of peripheral arterial disease in a defined population. Circulation 71:510–515
5. Hiatt WR, Hirsch AT, Regensteiner JG, Brass EP; Vascular Clinical Trialists (1995) Clinical trials for claudication: assessment of exercise performance, functional status and clinical endpoints. Circulation 92:614–621
6. Hiatt AT, Criqui MH, Treat-Johnson D et al (2001) Peripheral arterial disease detection, awareness and the treatment in primary care. JAMA 286:1317–1324
7. Smith GD, Shipley MJ, Rose G (1990) Intermittent claudication, heart disease risk factors and mortality: the Whitehall study. Circulation 82:1925–1931
8. Leng GC, Fowkes FGR (1993) The epidemiology of peripheral vascular disease. Vasc Med Rev 4:5–18
9. Imparato AM, Kim GE, Davidson T et al (1975) Intermittent claudication: its natural course. Surgery 78:795–799
10. Cronenwett JL, Warner KG, Davidson T et al (1984) Intermittent claudication – current results of non-operative management. Arch Surg 119:430–435
11. Walsh DB, Gilbertson JJ, Zwolak RM et al (1991) The natural history of superficial femoral artery stenoses. J Vasc Surg 14:299–304
12. McAllister FF (1976) The fate of patients with intermittent claudication managed non-operatively. Am J Surg 132:593–595
13. Fowkes FG, Allen PL, Tsampoulas C et al (1992) Validity of duplex scanning in the detection of peripheral arterial disease in the general population. Eur J Vasc Surg 6:31–35

14. Criqui MH, Fronek A, Klauber MR, Barrett-Connor E, Gabriel S (1985) The sensitivity, specificity and predictive value of traditional clinical evaluation of peripheral arterial disease: results from noninvasive testing in a defined population. Circulation 71:516–522
15. Criqui MH, Langer RD, Fronek A, Feigelson HS, Klauber MR, McCann TJ et al (1992) Mortality over a period of 10 years in patients with peripheral arterial disease. N Engl J Med 325:381–386
16. Vogt MT, McKenna M, Anderson SJ et al (1993) Prevalence and correlates of lower extremity arterial disease in elderly woman. Am J Epidemiol 137:559–568
17. Newman AB, Siscovick DS, Manolio TA et al (1993) Ankle arm index as a marker of atherosclerosis in Cardiovascular Health Study (CHS) Collaborative Research Group. Circulation 88:837–845
18. Newman AB, Sutton-Tyrrrell K, Vogt MT, Kuller LH (1993) Morbidity and mortality in hypertensive adults with a low ankle/arm blood pressure index. JAMA 270:487–489
19. Applegate WB (1993) Ankle/arm pressure index: a useful test for clinical practice. JAMA 270:497–498
20. Wensing PJ (1995) Arterial Tortuosity in the femoropopliteal region during knee flexion. J Anat 186:133–139
21. Jaffe MR (2004) The nature of SFA disease. Endovascular Today 3:3–5
22. Drisko K (2004) Characterizing the unique dynamics of the SFA. Endovascular Today 7(Supplement):6–8
23. Nesbitt E, Schmidt-Trucksass A, Il'yasov KA, Weber H, Huonker M, Laubenberger J et al (2000) Assessment of arterial blood flow characteristics in normal and atherosclerotic vessels with the fast Fourier flow method. MAGMA 10(1):27–34
24. Crawford DW, Barndt RJ, Back LH (1976) Surface characteristics of normal and atherosclerotic human arteries including observations suggesting interaction between flow and intimal morphology. Lab Invest 34(5):463–470
25. Strandness DE (1993) Duplex scanning in vascular disorders, 2nd ed. Raven Press, New York
26. Nerem RM (1992) Vascular fluid mechanics, the arterial wall, and atherosclerosis. J Biomech Eng 114(3):274–282
27. Davies PF, Polacek DC, Shi C, Helmke BP (2002) The convergence of haemodynamics, genomics, and endothelial structure in studies of the focal origin of atherosclerosis. Biorheology 39(3–4):299–306
28. Stary HC, Bleakley-Chandler A, Dinsmore RE, Fuster V, Glagov S, Insull WJ et al (1995) A definition of advanced types pf atherosclerotic lesions and a histological classification of atherosclerosis: a report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. Circulation 92:1355–1374
29. Stary HC, Chandler AB, Glagov S, Guyton JR, Insuli WJ, Rosenfeld ME et al (1994) A definition of initial fatty streak and intermediate lesions of atherosclerosis: a report from the Committee on Vascular Lesions of the Council on Atherosclerosis, American Heart Association. Arterioscler Thromb 14:840–856
30. Schenk EA (1973) Pathology of occlusive disease of the lower extremities. Cardiovasc Clin 5(1):287–310
31. Lee RT, Libby P (1997) The unstable atheroma. Arterioscler Thromb Vasc Biol 17(10): 1859–1867
32. Libby P (2001) Current concepts of the pathogenesis of the acute coronary syndromes. Circulation 104(3):365–372
33. Aikawa M, Libby P (2004) The vulnerable atherosclerotic plaque: pathogenesis and therapeutic approach. Cardiovasc Pathol 13(3):125–138
34. Lee AJ, Fowkes FG, Rattray A, Rumley A, Lowe GD (1996) Hemostatic and rheological factors in intermittent claudication: the influence of smoking and extent of arterial disease. Br J Haematol 92(1):226–230
35. TASC II Working Group (2007) TASC-II: Inter-Society Consensus for the Management of Peripheral Vascular Arterial Disease. J Vasc Surg 45(suppl S):S1–S67
36. TASC Working Group (2000) Management of peripheral arterial disease (PAD): Transatlantic Inter-Society Consensus (TASC). Eur J Vasc Endovasc Surg 19:S1–S244

37. Perlman H, Maillard L, Krasinski K, Walsh K (1997) Evidence for the rapid onset of apoptosis in medial smooth muscle cells after balloon injury. Circulation 95:981–987
38. Clowes AW, Reidy MA, Clowes MM (1983) Kinetics of cellular proliferation after arterial injury. I: Smooth muscle cell growth in the absence of endothelium. Lab Invest 49:327–333
39. Clowes AW, Schwartz SM (1985) Significance of quiescent smooth muscle cell migration in the injured rat carotid artery. Circ Res 56:139–145
40. Hanke H, Strohschneider T, Oberhoff M, Betz E, Karsch KR (1990) Time course of smooth muscle cell proliferation in the intima and media of arteries following experimental angioplasty. Circ Res 67:651–659
41. Majesky MW, Schwartz SM, Clowes MM, Clowes AW (1987) Heparin regulates smooth muscle S phase entry in the injured rat carotid artery. Circ Res 61:296–300
42. More RS, Rutty G, Underwood MJ, Brack MJ, Gershlick AH (1994) Assessment of myointimal cellular kinetics in a model of angioplasty by means of proliferating cell nuclear antigen expression. Am Heart J 128:681–686
43. Waksman R, Rodriquez JC, Robinson KA, Cipolla GD, Crocker IR, Scott NA et al (1997) Effect of intravascular irradiation on cell proliferation, apoptosis and vascular remodelling after balloon overstretch injury of porcine coronary arteries. Circulation 96:1944–1952
44. Hehrlein C, Gollan C, Donges K, Metz J, Riessen R, Fehsenfeld P et al (1995) Low-dose radioactive endovascular stents prevent smooth muscle cell proliferation and neointimal hyperplasia in rabbits. Circulation 92(6):1570–1575
45. Bauriedel G, Skowasch D, Jabs A, Dinkelbach S, Andrie R, Schiele TM et al (2002) Insights into vascular pathology after intracoronary brachytherapy. Z Kardiol 91(suppl 3):1–9
46. Schwartz SM (1997) Smooth muscle migration in atherosclerosis and restenosis. J Clin Invest 100(11 suppl):S87–S89
47. Davies MG, Hagen P-O (1994) Pathobiology of intimal hyperplasia. Br J Surg 81:1254–1269
48. Scott NA, Martin F, Simonet L, Dunn B, Ross CE, Wilcox JN (1995) Contribution of adventitial myofibroblasts to vascular remodelling and lesion formation after experimental angioplasty in pig coronaries (abstract). FASEB J 9:A845
49. Ferrer P, Valentine M, Jenkins-West T, Gale T, Gu K, Havens C et al (1996) Periadventitial changes in the balloon injured rat crotid artery (abstract). FASEB J 10:A618
50. Sarembock IJ, LaVeau PJ, Sigal SL, Timms I, Sussman J, Haudenschild C et al (1989) Influence of inflation pressure and balloon size on the development of intimal hyperplasia after balloon angioplasty. A study in the atherosclerotic rabbit. Circulation 80:1029–1040
51. Clowes AW, Clowes MM, Fingerle J, Reidy MA (1989) Kinetics of cellular proliferation after arterial injury. V: Role of acute distension in the induction of smooth muscle cell proliferation. Lab Invest 60:360–364
52. Andersen HR, Maeng M, Thorwest M, Falk E (1996) Remodeling rather than neointimal formation explains luminal narrowing after deep vessel wall injury – insights from a porcine coronary (re)stenosis model. Circulation 93:1716–1724
53. Moornekamp FN, Borst C, Post MJ (1996) Endothelial cell recoverage and intimal hyperplasia after endothelium removal with or without smooth muscle cell necrosis in the rabbit carotid artery. J Vasc Res 33:146–155
54. Cwikiel W (2002) Restenosis after balloon angioplasty and/or stent insertion – origin and prevention. Acta Radiologica 43:442–454
55. Lowe HC, Oesterle SN, Khachigan LM (2002) Coronary in stent restenosis: current status and future strategies. J Am Coll Cardiol 39(2):183–193
56. Mehran R, Dangas G, Abizaid A et al (1999) Angiographic patterns of in stent restenosis: classification and implications for longterm outcome. Circulation 100:1872–1878
57. Hoffman R, Mintz GS, Dussaillant RG, Popma JJ, Pichard AD, Satler LF et al (1996) Patterns and mechanisms of in stent restenosis: a serial intravascular ultrasound study. Circulation 94:1247–1254
58. Moreno PR, Palacios IF, Leon MN, Rhodes J, Fuster V, Fallon JT (1999) Histopathologic comparison of human coronary in stent and post balloon angioplasty restenotic tissue. Am J Cardiol 84:462–466

59. Virmani R, Farb A (1999) Pathology of instent restenosis. Curr Opin Lipidol 10:499–506
60. Baier RE, Dutton RC (1969) Initial events in interaction of blood with a foreign surface. J Biomed Mater Res 3:191
61. Emneus H, Stenram U (1965) Metal implants in the human body. Acta Orthop Scand 36:116
62. Parsson H, Cwikiel W, Johansson K, Swartbol P, Norgren L (1994) Deposition of platelets and neitrophils on porcine iliac arteries and angioplasty and Wallstent placement compared with angioplasty alone. Cardiovasc Intervent Radiol 17:190
63. Bai H, Masuda J, Sawa Y et al (1994) Neointima formation after vascular stent implantation: spatial and chronological distribution of smooth muscle cell proliferation and phenotypic modulation. Arterioscler Thromb 14:1846
64. Schwartz RS, Huber KC, Murphy JG et al (1992) Restenosis and the proportional neointimal response to coronary artery injury: results in a porcine model. J Am Coll Cardiol 19:267–274
65. Kornowski R, Hong MK, Fermin OT, Bramwell O, Wu H, Leon MB (1998) In-stent restenosis: contributions of inflammatory responses and arterial injury to neointimal hyperplasia. J Am Coll Cardiol 31:224–230
66. Sanada JL, Matsui O, Yoshikawa J, Matsuoka T (1998) An experimental study of endovascular stenting with special reference to the effects on the aortic vasa vasorum. Cardiovasc Intervent Radiol 21:45
67. Ramaswami G, Tegos T, Nicolaides AN, Dhanjil S, Griffin M, Al-Kutoubi A et al (1999) Ultrasonic plaque character and outcome after lower limb angioplasty. J Vasc Surg 29(1): 110–121
68. Marks NA, Ascher E, Hingorani AP, Shiferson A, Puggioni A (2008) Gray-scale median of the atherosclerotic plaque can predict success of lumen re-entry during subintimal femoral-popliteal angioplasty. J Vasc Surg 47(1):109–116
69. Met R, Bipat S, Legemate DA, Reekers JA, Koelemay MJ (2009) Diagnostic performance of computed tomography angiography in peripheral arterial disease: a systematic review and meta-analysis. JAMA 301(4):415–424
70. Heijenbrok-Kal MH, Kock MC, Hunink MG (2007) Lower extremity arterial disease: multidetector CT angiography meta-analysis. Radiology 245(2):433–439
71. Mell M, Tefera G, Thornton F, Siepman D, Turnipseed W (2007) Clinical utility of time-resolved imaging of contrast kinetics (TRICKS) magnetic resonance angiography for infrageniculate arterial occlusive disease. J Vasc Surg 45(3):543–548
72. Huegli RW, Thalhammer C, Jacob AL, Jaeger K, Bilecen D (2008) Intra-arterial MR-angiography on an open-bore MR-scanner compared to digital-subtraction angiography of the infra-popliteal runoff in patients with peripheral arterial occlusive disease. Eur J Radiol 66(3):519–525
73. Collins R, Cranny G, Burch J, Aguiar-Ibáñez R, Craig D, Wright K et al (2007) A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease. Health Technol Assess 11(20):iii–iv, xi–xiii, 1–184
74. Kashyap VS, Pavkov ML, Bishop PD, Nassoiy SP, Eagleton MJ, Clair DG et al (2008) Angiography underestimates peripheral atherosclerosis: lumenography revisited. J Endovasc Ther 15(1):117–125
75. Bishop PD, Feiten LE, Ouriel K, Nassoiy SP, Pavkov ML, Clair D et al (2008) Arterial calcification increases in distal arteries in patients with peripheral arterial disease. Ann Vasc Surg 22(6):799–805
76. Gussenhoven EJ, vanderLugt A, Pasterkamp G, vanderBerg FG, Sie LH, Vischjager M et al (1995) Intravascular ultrasound predictors of outcome after peripheral balloon angioplasty. Eur J Vasc Endovasc Surg 10(3):279–288
77. Vogt KC, Just S, Rasmussen JG, Schroeder TV (1997) Prediction of outcome after femoropopliteal balloon angioplasty by IVUS. Eur J Vasc Endovasc Surg 13(6):563–568
78. vanderLugt A, Gussenhoven EJ, Pasterkamp G, Stijnen T, Reekers JA, vanderBerg FG et al (1998) Intravascular ultrasound predictor of restenosis after balloon angioplasty of the femoropopliteal artery. Eur J Vasc Endovasc Surg 16(2):110–119

79. Spijkerboer AM, Nass PC, deValois JC, vanderGraaf Y, Eikelboom BC, Mali WP (1996) Evaluation of femoropopliteal arteries with duplex ultrasound after angioplasty, Can we predict results at one year. Eur J Vasc Endovasc Surg 12(4):418–423
80. van Lankeren W, Gussenhoven EJ, Honkoop J, Stijnen T, van Overhagen H, Wittens CHA et al (1999) Plaque area increase and vascular remodeling contribute to lumen area change after PTA of the femoropopliteal artery: an intravascular ultrasound study. J Vasc Surg 29(3): 430–441
81. Capek P, McLean GK, Berkowitz HD (1991) Femoropopliteal angioplasty. Factors influencing long-term success. Circulation 83(2 suppl):170–180
82. Hunink MG, Donaldson MC, Meyerovitz MF, Polak JF, Whittemore AD, Kandarpa K et al (1993) Risks and benefits of femoropopliteal percutaneous balloon angioplasty. J Vasc Surg 17:183–192
83. Wilson SE, Wold GL, Cross AP (1989) Percutaneous transluminal angioplasty versus operation for peripheral atherosclerosis: report of a prospective randomized trial in a selected group of patients. J Vasc Surg 9(1):1–9
84. Treiman GS, Ichikawa L, Treiman RL, Cohen JL, Cossman DV, Wagner WH et al (1994) Treatment of recurrent femoral or popliteal artery stenosis after percutaneous transluminal angioplasty. J Vasc Surg 20:577–587
85. Milford MA, Weaver FA, Lundell CJ et al (1988) Femoropopliteal PTA for limb salvage. J Vasc Surg 8:292–299
86. Blair JM, Gewertz BL, Moosa H, Lu C-T, Zarins CK (1989) PTA versus surgery for limb threatening ischemia. J Vasc Surg 9:698–703
87. Armstrong MW, Torrie EP, Galland RB (1992) Consequences of immediate failure of percutaneous transluminal angioplasty. Ann R Coll Surg Engl 74(4):265–268
88. Galaria II, Surowiec SM, Rhodes JM, Illig KA, Shortell CK, Davies MG (2005) Implications of early failure of superficial femoral artery endoluminal interventions. Ann Vasc Surg 19(6):787–792
89. Tatò F, Hoffmann U, Weber C, Reiser M, Rieger J (2006) Comparison of angiography. Duplex sonography and intravascular ultrasound for the graduation of femoropopliteal stenoses before and after balloon angioplasty. Ultrasound Med Biol 32(12):1837–1843
90. Dotter CT, Judkins MP (1964) Transluminal treatment of arteriosclerotic obstruction. Circulation 30:654–670
91. Gruntzig AR (1974) Percutane Rekanalization chronischer Arterieller Verschlusse mit einem neuen Dilatationskatheter. Duetsche Medizinische Wochenschrift 99:2502–2510
92. Matsi PJ, Manninen HI, Soder HK, Mustonen P, Kouri J (1995) PTA in femoral artery occlusions: primary and longterm results in 107 claudicant patients using femoral and popliteal catheterization techniques. Clin Radiol 50(4):237–244
93. Reekers JA, Kromhout JG, Jacobs MJ (1994) Percutaneous intentional extraluminal recanalization of the femoropopliteal artery. Eur J Vasc Surg 8(6):723–728
94. Bolia A, Miles KA, Brennan J, Bell PR (1990) Percutaneous transluminal angioplasty of occlusions of the femoral and popliteal arteries by subintimal dissection. Cardiovasc Intervent Radiol 13(6):357–363
95. Antusevas A, Aleksynas N, Kaupas RS, Inciura D, Kinduris S (2008) Comparison of results of subintimal angioplasty and percutaneous transluminal angioplasty in superficial femoral artery occlusions. Eur J Vasc Endovasc Surg 36(1):101–106
96. Yilmaz S, Sindel T, Ceken K, Alimoglu E, Luleci E (2001) Subintimal recanalization of long superficial femoral artery occlusions through the retrograde popliteal approach. Cardiovasc Intervent Radiol 24(3):154–160
97. Bakken AM, Palchik E, Hart JP, Rhodes JM, Saad WE, Davies MG (2007) Impact of diabetes on the outcomes of superficial femoral artery endoluminal interventions. J Vasc Surg 46: 946–958
98. Maca T, Schillinger M, Hamwi A, Mlekusch W, Sabeti S, Wagner O et al (2005) Insulin. C-peptide. and restenosis after femoral artery balloon angioplasty in type II diabetic and non-diabetic patients. J Vasc Interv Radiol 16(1):31–35

99. Becquemin J-P, Cavillon A, Haiduc F (1994) Surgical transluminal femoropopliteal angioplasty: multivariate analysis outcome. J Vasc Surg 19:495–502
100. Johnston KW (1992) Femoral and popliteal arteries: re-analysis of results of balloon angioplasty. Radiology 183:767–771
101. Lapeine P, Chabrot P, Chahid T, Ravel A, Boire JY, Garcier JM et al (2005) Femoral superficial artery angioplasty: long term results, initial predictive factors. 101 patients. J Mal Vasc 30(5):291–295
102. Exner M, Schillinger M, Minar E, Mlekusch W, Sabeti S, Endler G et al (2004) Interleukin-6 promoter genotype and restenosis after femoropopliteal balloon angioplasty: initial observations. Radiology 231(3):839–844
103. Wahlgren CM, Sten-Linder M, Egberg N, Kalin B, Blohmé L, Swedenborg J (2006) The role of coagulation and inflammation after angioplasty in patients with peripheral arterial disease. Cardiovasc Intervent Radiol 29(4):530–535
104. Ray B, Chetter IC, Lee HL, Ettelaie C, McCollum PT (2007) Plasma tissue factor is a predictor for restenosis after femoropopliteal angioplasty. Br J Surg 94(9):1092–1095
105. Ko YG, Kim JS, Choi DH, Jang Y, Shim WH (2007) Improved technical success and midterm patency with subintimal angioplasty compared to intraluminal angioplasty in long femoropopliteal occlusions. J Endovasc Ther 14(3):374–381
106. Scott EC, Biuckians A, Light RE, Scibelli CD, Milner TP, Meier G III et al (2007) Subintimal angioplasty for the treatment of claudication and critical limb ischemia: 3-year results. J Vasc Surg 46(5):959–964
107. Schanzer A, Owens CD, Conte MS, Belkin M (2007) Superficial femoral artery percutaneous intervention is an effective strategy to optimize inflow for distal origin bypass grafts. J Vasc Surg 45(4):740–743
108. Lantis J II, Jensen M, Benvenisty A, Mendes D, Gendics C, Todd G (2008) Outcomes of combined superficial femoral endovascular revascularization and popliteal to distal bypass for patients with tissue loss. Ann Vasc Surg 22(3):366–371
109. Heiss HW, Mathias K, Beck AH et al (1987) Efficacy of acetylsalicylic acid and dipyridamole for prevention of recurrence of femoral and popliteal arterial lesions following PTA. Cor Vasa 1:25–34
110. Minar E, Ehringer H, Ahmadi R, Dudczak R, Leitha T, Koppensteiner R et al (1989) Platelet deposition at angioplasty sites and its relation to restenoses in human iliac and femoropopliteal arteries. Radiology 170:767–772
111. Minar E, Ahmadi A, Koppensteiner R, Maca T, Stuempflen A, Ugurluoglu A et al (1995) Comparison of effects of high dose and low dose aspirin on restenosis after femoropopliteal PTA. Circulation 91:2167–2173
112. Ansel GM, Silver MJ, Botti CFJ, Rocha-Singh K, Bates MC, Rosenfield K et al (2006) Functional and clinical outcomes of nitinol stenting with and without abciximab for complex superficial femoral artery disease: a randomized trial. Catheter Cardiovasc Interv 67(2): 288–297
113. Yamawaki T, Yamada A, Fukumoto Y, Kishi T, Sobashima A, Kuwata K et al (2007) Statin therapy may prevent restenosis after successful coronary intervention, independent of lipid-lowering effect and CRP level. Fukuoka Igaku Zasshi 98(6):260–269
114. Hagenaars T, Gussenhoven EJ, Kranendonk SE, Blankensteijn JD, Honkoop J, van der Linden E et al (2000) Early experience with intravascular ultrasound in evaluating the effect of statins on femoropopliteal arterial disease: hypothesis-generating observations in humans. Cardiovasc Drugs Ther 14(6):635–641
115. Hagenaars T, Gussenhoven EJ, van Sambeek MR, Jukema JW, Kranendonk SE, Bom N (2000) Effect of simvastatin on restenosis after percutaneous transluminal angioplasty of femoropopliteal arterial obstruction. Am J Cardiol 86(7):774–776
116. Zdanowski Z, Albrechtsson U, Lundin A, Jonung T, Ribbe E, Thorne J et al (1999) Percutaneous transluminal angioplasty with or without stenting for femoropopliteal occlusions? A randomized controlled study. Int Angiol 18(4):251–255

117. Vroegindeweij D, Vos LD, Buth J, Bosch HC (1997) Balloon angioplasty combined with primary stenting versus balloon angioplasty alone in femoropopliteal obstructions: a comparative randomized study. Cardiovasc Intervent Radiol 20:420–425
118. Cejna M, Schoder M, Lammer J (1999) PTA versus stent in femoropopliteal obstruction. Radiologie 39(2):144–150
119. Becquemin J-P, Favre J-P, Marzelle J, Nemoz C, Corsin C, Leizorovicz A (2003) Systematic versus selective stent placement after SFA balloon angioplasty: a multicenter prospective randomized study. J Vasc Surg 37:487–494
120. Chatelard P, Guibort C (1996) Longterm results with a palmez stent in the femoropopliteal arteries. J Cardiovasc Surg 37(suppl 1):67–72
121. Sapoval M, Long AL, Raynaud AC, Beyssen BM, Flessinger JN, Gaux JC (1992) Femoropopliteal stent placement: longterm results. Radiology 184:833–839
122. Cheng SWK, Ting ACW, Wong J (2001) Endovascular stenting of SFA stenosis and occlusions: results and risk factor analysis Cardiovasc Surg 9(2):133–140
123. Liermann D, Bottcher HD, Kollath J et al (1994) Prophylactic endovascular radiotherapy to prevent intimal hyperplasia after stent implantation in femoropopliteal arteries. Cardiovasc Intervent Radiol 17:12–16
124. Strecker EP, Boos IB, Goettmann D (1997) Femoropoplitreal artery stent placement: evaluation of longterm success. Radiology 205:375–383
125. Rousseau H, Raillat C, Joffre F, Knight CJ, Ginestet MC (1989) Treatment of femoropopliteal stenoses by means of self-expandable endoprosthesis: mid term results. Radiology 172: 961–964
126. Zollikofer CL, Antonucci F, Pfyffer M et al (1991) Arterial stent placement with use of the Wallstent: midterm results of clinical experience. Radiology 179:449–456
127. Sabeti S, Mlekusch W, Amighi J, Minar E, Schillinger M (2005) Primary patency of long-segment self-expanding nitinol stents in the femoropopliteal arteries. J Endovasc Ther 12(1):6–12
128. Ihnat DM, Duong ST, Taylor ZC, Leon LR, Mills JL, Goshima KR et al (2008) Contemporary outcomes after superficial femoral artery angioplasty and stenting: the influence of TASC classification and runoff score. J Vasc Surg 47(5):967–974
129. Deutschmann HA, Schedibauer P, Berczi V, Portugaller H, Tauss J, Hausegger KA (2001) Placement of hemobahn stent-grafts in femoropopliteal arteries: early experience and midterm results in 18 patients. J Vasc Interv Radiol 12:943–949
130. Blum U, Fugel P, Ebert S et al (1996) Stentgraftiing for stenotic/occlusive lesions of femoropopliteal arteries. In: 38th Annual World Congress of the International College of Angiology, Cologne, Germany, 1996
131. Link J, Muller-Hulsbeck S, Brossmann J, Schwarzenberg H, Steffens JC, Heller M (1996) Unimantelte stents bei oberschenkelarterienverschlussen: ein erfahrungsbericht nach 2 jahrigeranwendung (abstract). ROFO 164:107
132. Henry M, Amor M, Cragg M et al (1996) Occlusive and aneurrysm peripheral arterial disease: assessment of a stent graft system. Radiology 201:717–724
133. Krankenberg H, Schlüter M, Steinkamp HJ, Bürgelin K, Scheinert D, Schulte KL et al (2007) Nitinol stent implantation versus percutaneous transluminal angioplasty in superficial femoral artery lesions up to 10 cm in length: the femoral artery stenting trial (FAST). Circulation 116(3):285–292
134. Zeller T, Tiefenbacher C, Steinkamp HJ, Langhoff R, Wittenberg G, Schlüter M et al (2008) Nitinol stent implantation in TASC A and B superficial femoral artery lesions: the Femoral Artery Conformexx Trial (FACT). J Endovasc Ther 15(4):390–398
135. Boeckler D, Blaurock P, Mannsman U, Schwarzbach M, Seelos R, Schumacher H et al (2005) Early surgical outcomes after failed primary stenting for lower limb occlusive disease. J Endovasc Ther 12(1):13–21
136. Mwipatayi BP, Hockings A, Hofmann M, Garbowski M, Sieunarine K (2008) Balloon angioplasty compared with stenting for treatment of femoropopliteal occlusive disease: a meta-analysis. J Vasc Surg 47(2):461–469

137. He N EY, Wang Y, Fan H (2008) Percutaneous transluminal angioplasty (PTA) alone versus PTA with balloon-expandable stent placement for short-segment femoropopliteal artery disease: a metaanalysis of randomized trials. J Vasc Interv Radiol 19(4):499–503
138. Duda SH, Bosiers M, Lammer J, Scheinert D, Zeller T, Tielbeek A (2005) The SIROCCO II trial: sirolimus-eluting versus bare nitinol stent for obstructive superficial femoral artery disease. J Vasc Interv Radiol 16:331–338
139. Duda SH, Poerner TC, Wiesinger B et al (2003) Drug eluting stents: potential applications for peripheral arterial occlusive disease. J Vasc Interv Radiol 14:291–301
140. Duda SH, Pusich B, Richter G et al (2002) Sirolimus-eluting stents for the treatment of obstructive superficial femoral artery disease, six month results. Circulation 106:1505–1509
141. Duda SH, Bosiers M, Lammer J, Scheinert D, Zeller T, Oliva V et al (2006) Drug-eluting and bare nitinol stents for the treatment of atherosclerotic lesions in the superficial femoral artery: long-term results from the SIROCCO trial. J Endovasc Ther 13(6):701–710
142. Cragg AH, Dake MD (1993) Percutaneous femoropopliteal graft placement. Radiology 187:643–648
143. Saxon RR, Coffman JM, Gooding JM, Natuzzi E, Ponec DJ (2003) Longterm results of ePTFE stent-graft versus angioplasty in the femoropopliteal artery: single center experience from a prospective randomized trial. J Vasc Interv Radiol 14:303–311
144. Wiesinger B, Beregi JP, Oliva VL, Dietrich T, Tepe G, Bosiers M et al (2005) PTFE-covered self-expanding nitinol stents for the treatment of severe iliac and femoral artery stenoses and occlusions: final results from a prospective study. J Endovasc Ther 12(2):240–246
145. Saxon RR, Coffman JM, Gooding JM, Ponec DJ (2007) Long-term patency and clinical outcome of the Viabahn stent-graft for femoropopliteal artery obstructions. J Vasc Interv Radiol 8(11):1341–1349
146. Alimi YS, Hakam Z, Hartung O, Boufi M, Barthèlemy P, Aissi K et al (2008) Efficacy of Viabahn in the treatment of severe superficial femoral artery lesions: which factors influence long-term patency? Eur J Vasc Endovasc Surg 35(3):346–352
147. Hartung O, Otero A, Dubuc M, Boufi M, Barthelemy P, Aissi K et al (2005) Efficacy of Hemobahn in the treatment of superficial femoral artery lesions in patients with acute or critical ischemia: a comparative study with claudicants. Eur J Vasc Endovasc Surg 30(3): 300–306
148. Lenti M, Cieri E, Rango P, Pozzilli P, Coscarella C, Bertoglio C et al (2007) Endovascular treatment of long lesions of the superficial femoral artery: results from a multicenter registry of a spiral, covered PTFE stent. J Vasc Surg 45(1):32–39
149. Savader SJ, Venbrux AC, Mitchell SE, Trerotola SO, Wang MC, Sneed TA et al (1994) Percutaneous transluminal atherectomy of the superficial and femoral and popliteal arteries: long-term results in 48 patients. Cardiovasc Intervent Radiol 17(6):312–318
150. Nelson PR, Powell RJ, Proia RR, Schermerborn ML, Fillinger MF, Zwolak RM et al (2001) Results of endovascular superficial femoral endarterectomy. J Vasc Surg 34(3):526–531
151. Wildenhain PM, Wholey MH, Jarmolowski CR, Hill KL (1994) Infrainguinal directional atherectomy: long term followup and comparison with percutaneous transluminal angioplasty. Cardiovasc Intervent Radiol 17(6):305–311
152. Moussa I, Moses J, DiMario C et al (1998) Stenting after optimal lesion debulking (SOLD) registry. Angiographic and clinical outcome. Circulation 98:1604–1609
153. Bramucci E, Angoli L, Merlini PA et al (1998) Adjunctive stent implantation following directional coronary atherectomy in patients with coronary artery disease. J Am Coll Cardiol 32:1855–1860
154. Dauerman H, Balm D, Cutlip D et al (1998) Mechanical debulking versus balloon angioplasty for the treatment of diffuse in-stent restenosis. Am J Cardiol 82:277–284
155. Dorros G, Iyer S, Lewin R, Zaitoun R, Mathiak L, Olson K (1991) Angiographic follow up and clinical outcome of 126 patients after percutaneous directional atherectomy (Simpson Atherocath) for occlusive peripheral vascular disease. Cathet Cardiovasc Diagn 22(2):79–84
156. vonPolnitz A, Nerlich A, Berger H, Hofling B (1990) Percutaneous peripheral atherectomy: angiographic and clinical follow up of 60 patients. J Am Coll Cardiol 15(3):682–688

157. Kuffer G, Spengel FA, Hansen R, Pfluger T, Nathrath W (1990) Simpson's atherectomy of the peripheral arteries: early results and follow up. Rofo 153(1):61–67
158. Kolvenbach R, Strosche H (1998) Longterm results after rotation angioplasty and catheter atherectomy. A retrospective analysis. J Cardiovasc Surg 39(1):15–18
159. Tielbeek AV, Vroegindeweij D, Buth J, Ladman GH (1996) Comparison of balloon angioplasty and Simpson atherectomy for lesions in the femoropopliteal artery: angiographic and clinical results of a prospective randomized trial. J Vasc Interv Radiol 7(6):837–844
160. Nakamura S, Conroy RM, Gordon IL, Deutsch LS, Maheswaran B, Antone CS et al (1995) A randomized trial of transcutaneus extraction atherectomy in femoral arteries: intravascular ultrasound observations. J Clin Ultrasound 23(8):461–471
161. Gordon IL, Conroy RM, Tobis JM, Kohl C, Wilson SE (1994) Determinants of patency after percutaneous angioplasty and atherectomy of occluded SFA. Am J Surg 168:115–119
162. Sarac TP, Altinel O, Bannazadeh M, Kashyap VS, Lyden SP, Clair D (2008) Midterm outcome predictors for lower extremity atherectomy procedures. J Vasc Surg 48(4):885–890
163. Chung SW, Sharafuddin MJ, Chigurupati R, Hoballah JJ (2008) Midterm patency following atherectomy for infrainguinal occlusive disease: a word of caution. Ann Vasc Surg 22(3): 358–365
164. Mintz GS, Kovach JA, Javier SP, Pichard AD, Kent KM, Popma JJ et al (1995) Mechanisms of lumen enlargement after excimer laser coronary angioplasty: an intravascular ultrasound study. Circulation 92(15):3408–3414
165. Topaz O, Das T, Dahm J, Madyhoon H, Perin E, Ebersole D (2001) Excimer laser revascularization: current indications, applications and techniques. Lasers Med Sci 16(2):72–77
166. Lammer J, Pilger E, Decrinis M et al (1992) Pulsed excimer laser versus continuous wave Nd:YAG laser versus conventional angioplasty of peripheral arterial occlusions: prospective controlled randomized trial. Lancet 340:1183–1188
167. Belli AM, Cumberland DC, Procter AE et al (1991) Total peripheral artery occlusions: conventional versus laser thermal recanalization with a hybrid probe in percutaneous angioplasty – results of a randomized trial. Radiology 181:57–60
168. Scheinert D, Laird JR, Schroeder M, Steinkamp H, Balzer JO, Biamino G (2001) Excimer laser-assisted recanalization of long chronic superficial femoral artery occlusion. J Endovasc Ther 8:156–166
169. Odink HF, deValois HC, Eikelboom BC (1995) Femoropopliteal artery recanalization: factors affecting clinical outcome of conventional and laser assisted balloon angioplasty. Cardiovasc Intervent Radiol 18(3):162–167
170. Steinkamp HJ, Rademaker J, Wissgott C, Scheinert D, Werk M, Settmacher U et al (2002) Percutaneous transluminal laser angioplasty versus balloon dilation for treatment of popliteal artery occlusions. J Endovasc Ther 9(6):882–888
171. Rastan A, Sixt S, Schwarzwälder U, Kerker W, Bürgelin K, Frank U et al (2007) Initial experience with directed laser atherectomy using the CLiRpath photoablation atherectomy system and bias sheath in superficial femoral artery lesions. J Endovasc Ther 14(3):365–373
172. Nicholson T (1998) Percutaneous transluminal angioplasty and enclosed thrombolysis versus PTA in the treatment of femoropopliteal occlusion: results of a prospective randomized trial. Cardiovasc Intervent Radiol 21(6):470–474
173. Wyttenbach R, Gallino A, Alerci M, Mahler F, Cozzi L, DiValentino M et al (2004) Effects of percutaneous transluminal angioplasty and endovascular brachytherapy on vascular remodeling of human femoropopliteal artery by noninvasive magnetic resonance imaging. Circulation 110:1156–1161
174. Hagenaars T, Po IFLA, van Sambeek RHM, Coen VLMA, van Tongeren RBM, Gescher FM et al (2002) Gamma radiation induces positive vascular remodeling after balloon angioplasty: a prospective randomized IVUS scan study. J Vasc Surg 36(2):318–324
175. Minar E, Pokrajac B, Maca T, Ahmadi R, Fellner C, Mittlboeck M et al (2000) Endovascular brachytherapy for prophylaxis of restenosis after femoropopliteal angioplasty. Circulation 102:2694–2699

176. Krueger K, Landwehr P, Bendel M, Nolte M, Stuetzer H, Bongartz R et al (2002) Endovascular gamma irradiation of femoropopliteal de novo stenoses immediately after PTA: interim results of prospective randomized controlled trial. Radiology 224:519–528
177. Wolfram RM, Pokrajac B, Ahmadi R, Fellner C, Gyoengyoesi M, Haumer M et al (2001) Endovascular brachytherapy for prophylaxis against restenosis after long segment femoropopliteal placement of stents: initial results. Radiology 220:724–729
178. Waksman R, Laird JR, Jurkovitz CT et al (2001) Intravascular radiation therapy after balloon angioplasty of narrowed femoropopliteal arteries to prevent restenosis: results of the PARIS feasibility clinical trial. J Vasc Intervent Radiol 12:915–921
179. Pokrajac B, Pötter R, Maca T, Fellner C, Mittlböck M, Ahmadi R et al (2000) Intraarterial (192)Ir high-dose-rate brachytherapy for prophylaxis of restenosis after femoropopliteal percutaneous transluminal angioplasty: the prospective randomized Vienna-2-trial radiotherapy parameters and risk factors analysis. Int J Radiat Oncol Biol Phys 48(4):923–931
180. van Tongeren RB, van Sambeek MR, van Overhagen H, Coen VL, Schmitz PI, Gescher FM et al (2005) Endovascular brachytherapy for the prevention of restenosis after femoropopliteal angioplasty. Results of the VARA Trial. J Cardiovasc Surg (Torino) 46(4):437–443
181. Hansrani M, Overbeck K, Smout J, Stansby G (2002) Intravascular brachytherapy for peripheral vascular disease. Cochrane Database Syst Rev 4:CD003504
182. Wyttenbach R, Corti R, Alerci M, Cozzi L, Di Valentino M, Segatto JM et al (2007) Effects of percutaneous transluminal angioplasty and endovascular brachytherapy on vascular remodeling of human femoropopliteal artery: 2 years follow-up by noninvasive magnetic resonance imaging. Eur J Vasc Endovasc Surg 34(4):416–423
183. Cheema AN, Nili N, Li CW, Whittingham HA, Linde J, VanSuylen RJ et al (2003) Effects of intravascular cryotherapy on vessel wall repair in a balloon injured rabbit iliac artery model. Cardiovasc Res 59(1):222–233
184. Laird JR, Jaff MR, Biamino G, McNamara T, Scheinert D, Zetterlund P et al (2005) Femoropopliteal outcomes with cryoplasty: final results of the cryovascular safety registry. J Vasc Interv Radiol 16(8):1067–1073
185. Samson RH, Showalter DP, Lepore MRJ, Ames S (2006) CryoPlasty therapy of the superficial femoral and popliteal arteries: a single center experience. Vasc Endovascular Surg 40(6): 446–450
186. Fava M, Loyola S, Polydorou A, Papapavlou P, Polydorou A, Mendiz O et al (2004) Cryoplasty for femoropopliteal arterial disease: late angiographic results of initial human experience. J Vasc Interv Radiol 15(11):1239–1243
187. Bakken AM, Saad WEA, Davies MG (2008) Cryo balloon angioplasty broadens the role of primary angioplasty and reduces adjuvant stenting in complex superficial femoral artery (TASC C and D) lesion. J Am Coll Surg 206(3):524–532
188. Samson RH, Showalter DP, Lepore MJ, Nair DG, Merigliano K (2008) CryoPlasty therapy of the superficial femoral and popliteal arteries: a reappraisal after 44 months' experience. J Vasc Surg 48(3):634–637
189. McCaslin JE, Macdonald S, Stansby G (2007) Cryoplasty for peripheral vascular disease. Cochrane Database Syst Rev 17(4):CD005507
190. Canaud L, Alric P, Berthet JP, Marty-Ané C, Mercier G, Branchereau P (2008) Infrainguinal cutting balloon angioplasty in de novo arterial lesions. J Vasc Surg 48(5):1182–1188
191. Cotroneo AR, Pascali D, Iezzi R (2008) Cutting balloon versus conventional balloon angioplasty in short femoropopliteal arterial stenoses. J Endovasc Ther 15(3):283–291
192. Cutlip DE (2008) Stent thrombosis: historical perspectives and current trends. J Thromb Thrombolysis 10(1):89–101
193. Lam RC, Shah S, Faries PL, McKinsey JF, Kent KC, Morrissey NJ (2007) Incidence and clinical significance of distal embolization during percutaneous interventions involving the superficial femoral artery. J Vasc Surg 46(6):1155–1159
194. Al-Hamali S, Baskerville P, Fraser S, Walters H, Markus HS (1999) Detection of distal emboli in patients with peripheral arterial stenosis before and after iliac angioplasty: a prospective study. J Vasc Surg 29(2):345–351

195. Shammas NW, Dippel EJ, Coiner D, Shammas GA, Jerin M, Kumar A (2008) Preventing lower extremity distal embolization using embolic filter protection: results of the PROTECT registry. J Endovasc Ther 15(3):270–276
196. Suri R, Wholey MH, Postoak D, Hagino RT, Toursarkissian B (2006) Distal embolic protection during femoropopliteal atherectomy. Catheter Cardiovasc Interv 67(4):417–422
197. Karnabatidis D, Katsanos K, Kagadis GC, Ravazoula P, Diamantopoulos A, Nikiforidis GC et al (2006) Distal embolism during percutaneous revascularization of infra-aortic arterial occlusive disease: an underestimated phenomenon. J Endovasc Ther 13(3):269–280
198. Kamiya C, Sakamoto S, Tamori Y, Yoshimuta T, Higashi M, Tanaka R et al (2008) Long-term outcome after percutaneous peripheral intervention vs medical treatment for patients with superficial femoral artery occlusive disease. Circ J 72(5):734–739
199. Spronk S, Bosch JL, den Hoed PT, Veen HF, Pattynama PM, Hunink MG (2009) Intermittent claudication: clinical effectiveness of endovascular revascularization versus supervised hospital-based exercise training – randomized controlled trial. Radiology 250(2):586–595
200. Dosluoglu HH, Cherr GS, Lall P, Harris LM, Dryjski ML (2008) Stenting vs above knee polytetrafluoroethylene bypass for TransAtlantic Inter-Society Consensus-II C and D superficial femoral artery disease. J Vasc Surg 48(5):1166–1174
201. Sabeti S, Czerwenka-Wenkstetten A, Dick P, Schlager O, Amighi J, Mlekusch I et al (2007) Quality of life after balloon angioplasty versus stent implantation in the superficial femoral artery: findings from a randomized controlled trial. J Endovasc Ther 14(4):431–437

Part II
Image Processing and Diagnostics

Brain MRI Segmentation

Stéphanie Bricq, Christophe Collet, and Jean-Paul Armspach

Introduction

Segmentation is an important step for quantitative analysis of brain images and for the study of many brain disorders. Indeed, structural changes in the brain can be due to some brain disorders. The quantification of these changes, by measuring volumes of structures of interest, can be used to characterize disease severity or evolution. For instance, in morphometry, brain tissue segmentation enables to compare tissue volumes and to follow the evolution of some brain disorders. Before these measurements can be carried out, the labeling process must be performed. In this approach, the tissue types of interest are white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), but other approaches classify voxels according to their anatomical structure. As most studies involve large amounts of data, manual tracing of cerebral structures in MR images by a human expert is obviously a time-consuming process. Moreover, these manual segmentations are prone to large intra- and interobserver variability, which deteriorates the significance of the resulting segmentation analysis.

Thus, due to different artifacts appearing on MR images, segmentation is not obvious and remains a challenging task. The first point deals with spatial regularization required to overcome the disturbance added during the MRI formation. The second artifact that appears in MR images is the corruption by a bias field. The third important issue in MR segmentation is the partial volume effect (PVE). Due to the limited resolution of acquisition system, voxels along the boundaries are composed of two or more tissues. Therefore, it is necessary to consider this partial volume effect in order to achieve an accurate segmentation of brain tissues. In fact, hard segmentation (WM, GM, CSF) ignores this problem and therefore loses information concerning the tissue structure.

S. Bricq and C. Collet (✉)
Strasbourg University, Image Sciences, Computer Sciences and Remote Sensing Lab. (LSIIT UMR CNRS 7005) Strasbourg, France
e-mail: c.collet@unistra.fr

J.-P. Armspach
LINC-IPB, Faculté de Médecine, 4, rue Kirschleger F - 67085 STRASBOURG Cedex
e-mail: jparmspach@unistra.fr

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_3,

We propose an unsupervised framework for determining the amount of each substance in each voxel of the brain, including bias field correction, neighborhood information using the hidden Markov chain model (HMC) and using a probabilistic atlas as a complementary sensor. To carry out a regularized segmentation, we adopt the hidden Markov chain model. A bias field estimation step is added to correct intensity inhomogeneities. To achieve an accurate segmentation of brain tissues, partial volume effects are included in the model. Instead of assigning each voxel to a single tissue class (i.e., hard classification), the method we proposed estimates the proportion of each tissue type in each voxel without requiring any tunable parameter.

The atlas we used represents the a priori probability of a voxel being WM, GM, or CSF, after an image has been normalized to the same space. In this work, we assume that the registration of the images with the atlas has already been performed. Further information on registration methods can be found elsewhere (e.g., in our case, we use the method proposed in [30]).

Then, we extended this method to MS lesion detection. MS lesions are detected as voxels that are not well explained by a statistical model for normal brain MR images. To extract these outliers and to estimate the parameters of the different classes in a robust way, we used the trimmed likelihood estimator (TLE) [29].

Once the lesions have been detected, we focus on the refinement on the contours of the detected lesions. To have a more precise information on these contours, we propose to extend the framework described in [15, 27] for 2D images segmentation based on minimum description length (MDL) principle, to 3D brain MRI segmentation.

Brain MRI segmentation Based on Hidden Markov Chains

Hidden Markov Chain Modeling

Hilbert–Peano Scan

Thanks to the fractal Hilbert–Peano scan [6] (Fig. 1), hidden Markov chain models can be used to segment 3D data. The interest of Markov chain methods for image segmentation when compared with 3D Markov field models is that, being based on 1D modeling, they result in lower computing cost. Nevertheless, contrary to the hidden Markov field, in the Markov chain model the neighboring information is partially translated in the chain (through the 3D-Hilbert–Peano path): two neighbors in the chain are neighbors in the grid, but two neighbors in the cube can be far away in the chain. This scan is the best suited to take into account the concept of neighborhood. However, probably due to strong correlation within the data cube, this scan will weakly influence the segmentation results. The first step of segmentation algorithms based on Markov chains consists in transforming the image into a vector. Once all the processing has been carried out on the vector, we apply the inverse transformation on the segmented chain to obtain the final segmented image.

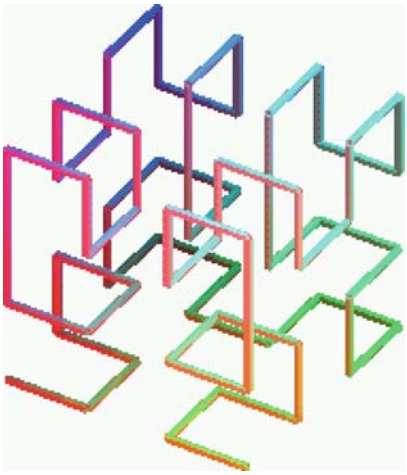


Fig. 1 Hilbert–Peano scans for 3D images

HMC Theory

Let us consider two sequences of random variables $X = (X_n)_{n \in S}$ the hidden process, and $Y = (Y_n)_{n \in S}$ the observed one, with S the finite set corresponding to the N voxels of an image. Y_n is a vector containing the observations from different modalities at voxel n: $Y_n = \{Y_n^1, \ldots, Y_n^l, \ldots, Y_n^L\}$ with L the number of modalities and Y_n^l the observations at modality l at voxel n. For multimodal images, we use multivariate normals with mean μ_k and covariance matrix Σ_k. Each X_n takes its value in a finite set of K classes $\Omega = \{\omega_1, \omega_2, \ldots, \omega_K\}$. Each Y_n^l takes its value in the set of real numbers $\mathbb{R}$.

X is a Markov chain if [31]:

$$P(X_{n+1} = \omega_{i_{n+1}} | X_n = \omega_{i_n}, \ldots, X_1 = \omega_{i_1}) = P(X_{n+1} = \omega_{i_{n+1}} | X_n = \omega_{i_n}) \tag{1}$$

Thus, X will be determined by two parameters:

- The initial distribution $\pi_i = P(X_1 = \omega_i)$
- The transition matrix $t_{ij}^n = P(X_{n+1} = \omega_j | X_n = \omega_i)$

Furthermore, we assume the homogeneity of the Markov chain, which means that the transition matrix is independent of the location n: $t_{ij}^n = t_{ij}$, for $1 \leq n < N$. Nevertheless, we do not assume the stationarity of the chain, which means that $\pi_i \neq P(X_n = \omega_i), \forall n \neq 1$. Figure 2 presents the HMC dependency graph.

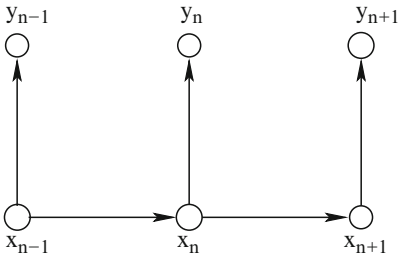


Fig. 2 Dependency graph of hidden Markov chain

Algorithm 1 presents the MPM segmentation based on a hidden Markov chain.

Algorithm 1 MPM segmentation based on a hidden Markov chain

1. Image vectorization using 3D Hilbert–Peano scan
2. Initialization
   a. For a priori parameters Φ_x

$$\pi_k = \frac{1}{K}, \forall k \tag{2}$$

$$t_{ij} = \begin{cases} \frac{3}{4} \text{ if } i = j \\ \frac{1}{4(K-1)} \text{ else} \end{cases} \tag{3}$$

   b. Data-driven parameters Φ_y are initialized using KMeans approach [9]
3. For each iteration
   a. Calculation of $\alpha_n(k)$ and $\beta_n(k)$
   b. Update of parameters $\Phi = \{\Phi_x, \Phi_y\}$
4. MPM segmentation

$$\hat{x}_n = \arg\max_{\omega_k} P(X_n = \omega_k | Y = y, \Phi) = \arg\max_{\omega_k} \alpha_n(k)\beta_n(k) \tag{4}$$

5. Inverse scan transformation to obtain the segmented image

Unifying Framework for Multimodal MRI Segmentation

In this section, we describe the unifying framework for unsupervised segmentation of multimodal brain MR images, including partial volume effect, bias field correction, information given by a probabilistic atlas, and neighborhood information based on HMC principles we presented in the previous section.

Probabilistic Atlas

For a Bayesian approach, the information brought by a probabilistic atlas may help and finally improve the accuracy of segmentation (Fig. 3). For this reason, a priori

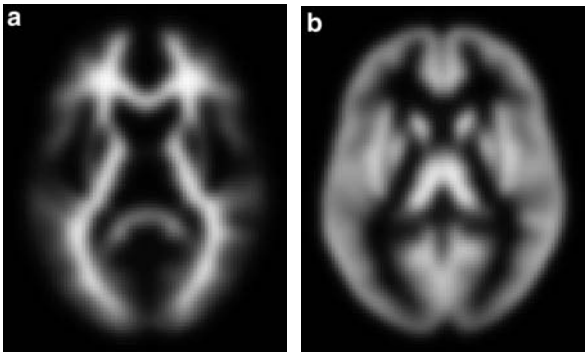


Fig. 3 Example of a priori probability images: (**a**) white matter atlas, (**b**) gray matter atlas

information was sometimes introduced in some models [2,5]. Ashburner and Friston proposed a modified maximum likelihood mixture model algorithm using probabilistic images for the segmentation [3,4]. The probability images they used, are the means of binary images of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) obtained from the segmentation of the original images. However, this method does not take into account neighborhood information, and therefore the quality of the segmentation is degraded if noise is present in the images. Other methods [26, 44] propose to use a digital brain atlas to initialize the iterative process. Each time, the main idea remains the same: the atlas is an interesting tool to drive the segmentation process. In [26], a brain atlas is used to segment the brain from nonbrain tissue and to compute prior probabilities for each class at each voxel location. In [2], a fuzzy atlas which indicates the probability distribution of each tissue type in the brain is used as a priori knowledge to correct misclassified voxels. In this last case, the neighborhood behavior is regularized thanks to a Markov random field modeling to be robust toward the noise.

The modeling scheme we proposed captures a priori information both in the neighborhood and in an atlas and in fact, this atlas is considered a complementary sensor. In this study, we use an atlas bringing a priori information deriving from averaged segmented MR images. In this way, the atlas provides probability maps of the three main tissues in the brain (WM, GM, CSF). $p(B_n|X_n = \omega_k) = b_n(k)$ denotes the tissue probability for class k at voxel n given by the atlas.

The model we proposed here is a hidden Markov chain with two independent sensors (Fig. 4). In this case, we have as information, both the intensity of the observed data and the probability maps. Thus, we have to find $\hat{x}_n = \arg\max_{\omega_k} P(X_n = \omega_k|Y, B)$, where $P(X_n = \omega_k|Y, B) = \alpha_n(k)\beta_n(k)$ with $\alpha_n = P(X_n = \omega_k|Y_{\leq n}, B_{\leq n})$ *forward* probability and $\beta_n = \frac{P(Y_{>n}, B_{>n}|X_n=\omega_k)}{P(Y_{>n}, B_{>n}|Y_{\leq n}, B_{\leq n})}$ *backward* probability.

With this link to the atlas, both a priori knowledge from the probability maps of each tissue classes and spatial organization will be considered for segmentation. The data driven parameters will depend on the observations and on the probability maps of the different tissue of the brain.

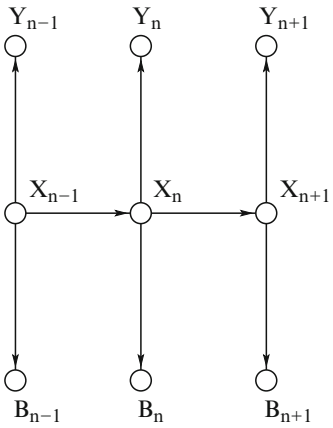


Fig. 4 Hidden Markov chain X with information given by a probabilistic atlas and the observations Y

Bias Field Correction

An artifact that appears in MR images is the corruption by a bias field. The most commonly used model is a multiplicative bias field with an additive noise:

$$y(x) = \tilde{y}(x)b(x) + n(x) \tag{5}$$

where $y(x)$ is the voxel intensity, $\tilde{y}(x)$ the "true" voxel intensity, $b(x)$ is a multiplicative bias field and $n(x)$ an additive Gaussian noise. In [7], Belaroussi et al. present an overview of existing methods for intensity nonuniformity correction. This bias field is frequently modeled as a linear combination of smooth basis functions. These functions are either spline functions [38] or polynomial basis functions [40, 43]. In [38], Sled et al. propose to correct nonuniformity by finding the smooth multiplicative field which maximizes the frequency content of $\mathbf{Z}$, where $\mathbf{Z}$ denotes the probability function of $\log \tilde{y}$. This method does not need a tissue model and can be applied without data segmentation. Other approaches interleave bias field estimation and tissue classification: bias field correction and segmentation are then included in the same framework by maximizing the same cost function. Most of these approaches are based on the EM algorithm [26, 43], estimating alternatively classification and bias field. In this way, different modelings for the bias field b are proposed. In [47], the bias field is modeled by a N-dimensional zero mean Gaussian prior probability density function, where Y_n is the observed MRI intensity at voxel n. Van Leemput et al. represent the bias field by a linear combination of polynomial basis functions [43]. Ashburner and Friston parameterize the bias field as an exponential of a linear combination of low frequency basis functions [5].

To take into consideration bias field inhomogeneities, a bias field estimation step is included to the processing chain we proposed. MRI intensity inhomogeneities are modeled by a bias field which multiplies the intensity data. Thanks to a logarithmic transformation of the intensities, the artifact can be modeled by an additive bias field [47]: $Z_n = \log Y_n$, where Y_n is the observed MRI intensity at voxel n.

We model the bias field as a linear combination $\sum_k c_k \psi_k(t)$ of smooth basis functions $\psi_k(t)$, as [43], whereas each class is modeled by a normal distribution.

$$p(Z_n = z_n | X_n = \omega_i) = \frac{1}{\sqrt{\det \Sigma_i}(2\pi)^{L/2}} \exp\left(-\frac{1}{2}\left[z_n - \mu_i - \sum_k c_k \psi_k(t_n)\right]^t {\Sigma_i}^{-1}\left[z_n - \mu_i - \sum_k c_k \psi_k(t_n)\right]\right) \quad (6)$$

Such an approximation is used in [22, 43, 47]. The algorithm presented in the previous section is slightly modified to include the bias field correction step. Polynomial basis functions were used as bias field basis functions.

Hidden Markov Chain with Partial Volume Effect

Instead of assigning each voxel to a single class, some authors estimate the amount of every tissue in each individual voxel. Indeed, due to the limited resolution of acquisition system, voxels along the boundaries between different classes are composed of two or more tissues. The classification of a voxel representing the main tissue type loses information about the tissue content.

In the literature, several papers have incorporated partial volume effect in segmentation: modified EM algorithms have been proposed to include the partial volume effect [18]. However, their method does not consider pixel neighborhood. In [8], Bosc et al. address issues related to partial volumes by modeling the acquisition process and using a high resolution label image, thereby obtaining high resolution segmentations from standard MR images.

Some approaches model the mixels with a spatial correlation between the proportions of the different tissue types. A Markovian a priori is introduced to impose this constraint: Choi et al. use a Markov random field (MRF) model to take into account spatial correlation between neighboring voxels [13] (this method needs multimodal images, which are not always available). Indeed, the probability of occurrence of each tissue type for a voxel is influenced by the tissue type of its neighbors. Some methods first classify the brain into different PV labels using a simple Potts model [37, 42]. Then, they assign to each voxel a tissue fraction for WM, GM and CSF based on its tissue label and its intensity in the MR image. However, this method requires one parameter to be chosen by the user: the β term of the Potts model that controls the strength of the prior is a user tunable parameter. Some methods only identify voxels containing PVE or do not try to estimate the amount of each tissue in each voxel: for example Ruan et al. propose to first classify voxels in pure tissues and mixture classes using an MRF prior [34]. The segmentation obtained is then used to assign each voxel to one single tissue type using a T1-specific feature, thus losing all notion of partial volume effect. In another paper, they use a fuzzy Markov random field model to take into account partial volume effect [35]. This method is unfortunately time-consuming due to the generation of the Markov field which is carried out using a Gibbs sampler. To solve

this problem, the estimation algorithm was performed in a region which is about $1/6$ of the entire volume, thereby losing information for computational reasons. In [45], Van Leemput et al. use a subvoxel label image L. The observed image $\tilde{Y} = \{\tilde{y}_i, i = 1, \ldots, I\}$ is considered the downsampling by a factor M of an ideal image $Y = \{y_i, i = 1, \ldots, J\}$ (with $J = I \times M$) without partial volume effect. Given the observed image $\tilde{Y}$, the aim is to reconstruct the subvoxel label L or rather to estimate the tissue fractions in each voxel. They use an EM algorithm [16] to estimate model parameters and missing data, i.e., unknown data L, Y from the observed one $\tilde{Y}$. They present three different prior models for the underlying label image L, especially an MRF model. This MRF model is defined on subvoxels, which constrain homogeneous regions of pure tissues bordered by PV voxels. However, this algorithm was implemented only for 2D images and requires about 20 min for one slice of an MR image.

The method we proposed estimates the PV fraction of pure tissue types in each voxel by adapting the HMC model presented in the previous section to include partial volume effects. Proposed method is a postprocessing step working on the nonlogtransformed data. During this step, the atlas is not used and the parameters of the bias field model are kept fixed. Let us denote by A_n the vector which contains proportions of each tissue class in voxel n.

$$A_n = \begin{bmatrix} a_n^1 \\ \vdots \\ a_n^{K-1} \end{bmatrix} \tag{7}$$

with $\forall n, \forall k, a_n^k$ denoting the proportion of tissue k in voxel n and taking its values within $[0, 1]$. Amounts of tissue K are deduced by: $a_n^K = 1 - \sum_{k=1}^{K-1} a_n^k$. a_n^k can take an infinity of values in $[0, 1]$, but in practice, we discretize it using a step of 0.1 or 0.2, which means that a_n^k takes its values in $\{0, 0.1, 0.2, 0.3, \ldots, 0.9, 1\}$ (respectively $\{0, 0.2, \ldots, 0.8, 1\}$).

We represent the intensity y of a voxel by a weighted sum of K pure tissues:

$$y = \sum_{i=1}^{K} a^i y_i^p \tag{8}$$

with $\sum_{i=1}^{K} a^i = 1$ and $a^i > 0$. a^i denotes the proportion of the pure class i and y_i^p a random variable representing the pure class i. We assume that each pure class is normally distributed $y_i^p \sim \mathcal{N}(\mu_i, \sigma_i^2)$ in intensity and that Y_i^p are independent. Therefore, for a given A: $Y|A \sim \mathcal{N}(\sum_{i=1}^{K} a^i \mu_i, \sum_{i=1}^{K} (a^i)^2 \sigma_i^2)$.

A is a Markov chain, thus we can apply the properties of Markov chains presented in previous section. A is determined by the initial distribution and the transition matrix. Figure 5 presents the dependency graph of an HMC with partial volume effect. We use the MPM [21] to obtain a labeling of the image.

$$\hat{A}_n = \arg\max_{A_n} P(A_n|Y) = \alpha_n \beta_n \tag{9}$$

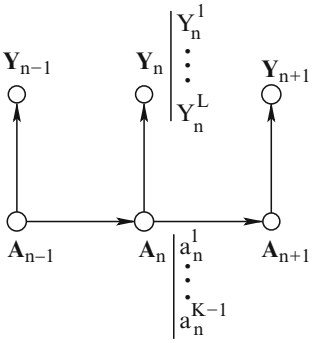


Fig. 5 Hidden Markov chain with partial volume effect

with $\alpha_n = P(A_n|Y_{\leq}n)$ *forward* probability and $\beta_n = \frac{P(Y_{>n}|A_n)}{P(Y_{>n}|Y_{\leq n})}$ *backward* probability. For each voxel, we compute the probabilities of every feasible label and pick the one with the highest probability.

Thus, we obtain K segmentation maps containing the amount of each tissue type in each voxel. As a conclusion, we proposed a new HMC model to take into account PVE, which is entirely unsupervised. The results we obtain (cf. following sections) show that contextual information can be correctly captured by the 3D Hilbert–Peano scan. The major advantage to use HMC stands for their ability to learn model parameters and to estimate a posteriori probabilities in exact manner on each site with low computing costs.

Algorithm Outline

We will assume that the registration of the images with the atlas [30] has already been performed. Therefore, in the following, we will not consider the registration step in computing time. Spatial normalization is performed using a nonrigid transformation described in [30]. The brain volume is then extracted by removing the skull using the brain extraction tool (BET) [39]. To initialize the *data driven* parameters $\Phi_y = \{\mu_k, \Sigma_k\}$, we use the K-Means algorithm [9]. Partial volume effects do not affect bias field correction. Thus, to speed up the algorithm, we first estimate the bias field before adding partial volume effects (see Fig. 6). We apply the HMC algorithm with bias field correction using a priori information contained in the probability maps. This leads to the estimation of the bias field and of μ and σ for pure classes. At this step of the algorithm, we could also provide a segmentation into three pure classes.

To reduce the numerical complexity, we only allow two different tissue types in a voxel at the same time. In fact, the probability of mixing more than two tissues is very low. Furthermore, the probability of the mixture of WM and CSF is low [34, 35, 37], which is why we enable only the WM/GM and the GM/CSF mixture in the brain. The last step of the algorithm consists in applying HMC segmentation with partial volume effects to obtain three maps containing the amounts of each pure tissue in each voxel (see on Fig. 6 the complete processing chain).

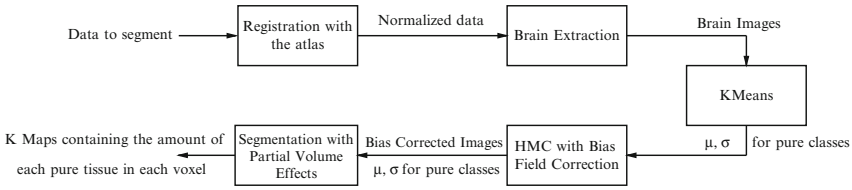


Fig. 6 Algorithm outline

Algorithm 2 presents the unifying framework we proposed to segment multimodal brain MRI.

Algorithm 2 Multimodal brain MRI segmentation based on a hidden Markov chain

1. Nonrigid registration of the images with the atlas [30]
2. Brain extraction using BET [39]
3. Image vectorization using 3D Hilbert–Peano scan [6]
4. Initialization

   a. For a priori parameters Φ_x

$$\pi_k = \frac{1}{K}, \forall k \tag{10}$$

$$t_{ij} = \begin{cases} \frac{3}{4} \text{ if } i = j \\ \frac{1}{4(K-1)} \text{ else} \end{cases} \tag{11}$$

   b. Data-driven parameters Φ_y are initialized using KMeans [9]
5. HMC with bias field correction (on log-transformed data), using atlas information. For each iteration

   a. Calculation of $\alpha_n(k)$ et $\beta_n(k)$
   b. Update of parameters Φ
6. HMC segmentation with partial volume estimation (on nonlog-transformed data), without using atlas information and keeping bias field parameters (estimated during the previous step) fixed.
7. MPM segmentation

$$\hat{x}_n = \arg\max_{\omega_k} P(X_n = \omega_k | Y = y, \Phi) = \arg\max_{\omega_k} \alpha_n(k)\beta_n(k) \tag{12}$$

8. Inverse Hilbert–Peano transformation to obtain the segmented image

Extension to MS Lesion Detection

Trimmed Likelihood Estimator

We detect MS lesions as outliers toward statistical model of normal brain images. To extract these outliers and to estimate the parameters of the different classes in a robust way, the trimmed likelihood estimator (TLE) was used. The TLE was

introduced in [29] and developed to estimate mixture of multivariate normals and generalized linear models in a robust way [46]. The main idea lies in removing the $n - h$ observations whose values would be highly unlikely to occur if the fitted model was true. The optimization scheme used to compute this estimator derives from the optimization scheme of the least trimmed squares (LTS) estimators of Rousseeuw and Leroy [33]. This algorithm was used to segment brain MRI by Aït-Ali in the frame of Gaussian mixtures without including neighborhood and atlas information [1].

The trimmed likelihood estimator (TLE) [23] is defined as:

$$\hat{\theta}_{\text{TLE}} = \arg \min_{\theta \in \Theta^p} \sum_{i=1}^{h} \psi(y_{\nu(i)}; \theta) \tag{13}$$

where for a fixed θ, $\psi(y_{\nu(1)}; \theta) \leq \psi(y_{\nu(2)}; \theta) \leq \cdots \leq \psi(y_{\nu(n)}; \theta)$, $\psi(y_i; \theta) = -\log f(y_i; \theta)$, $y_i \in \mathbb{R}^q$ for $i = 1, \ldots, n$ are i.i.d observations with probability density $f(y, \theta)$ depending on an unknown parameter $\theta \in \Theta^p \subset \mathbb{R}^p$. $\nu = (\nu(1), \ldots, \nu(n))$ is the corresponding permutation of the indices, which depends on θ, and h is the trimming parameter.

$$\hat{\theta}_{\text{TLE}} = \arg \max_{\theta \in \Theta^p} \prod_{i=1}^{h} f(y_{\nu(i)}; \theta) \tag{14}$$

General conditions for the existence of a solution of (13) are proved in [17]. Convergence and asymptotic properties are studied in [14].

FAST-TLE Algorithm

In [28], Neykov and Müller develop a fast iterative algorithm for derivation of the TLE. This FAST-TLE algorithm can be described as follows: given $H^{\text{old}} = \{y_{j_1}, \ldots, y_{j_n}\} \subset \{y_1, \ldots, y_n\}$,

- Compute $\hat{\theta}^{\text{old}} :=$ MLE (maximum likelihood estimator) based on H^{old}.
- Define $Q^{\text{old}} = \sum_{i=1}^{k} \psi(y_{j_i}, \hat{\theta}^{\text{old}})$.
- Sort $\psi(y_i, \hat{\theta}^{\text{old}})$ for $i = 1, \ldots, n$ in an ascending order: $\psi(y_{\nu(i)}, \hat{\theta}^{\text{old}}) \leq \psi(y_{\nu(i+1)}, \hat{\theta}^{\text{old}})$ and get the permutation $\nu = (\nu(1), \ldots, \nu(n))$.
- Define $H^{\text{new}} = \{y_{\nu(1)}, \ldots, y_{\nu(n)}\}$.
- Compute $\hat{\theta}^{\text{new}} :=$ MLE based on H^{new}.
- Define $Q^{\text{new}} = \sum_{i=1}^{k} \psi(y_{\nu(i)}, \hat{\theta}^{\text{new}})$.

Proposed Framework

To estimate parameters in a robust way and to detect lesions, we thus adapt the FAST-TLE algorithm presented in section "FAST-TLE Algorithm" to the HMC model. We use the following notations: $f_k(y_n; \theta) = P(Y_n = y_n | X_n = \omega_k)$ denotes

the likelihood of the observation y_n conditionally to $X_n = \omega_k$ and $b_n(k)$ represents the prior probability of voxel n to belong to class k given by the atlas B. This leads to:

1. Compute $\hat{\theta}^{(p-1)} :=$ MLE using the expectation-maximization (EM) algorithm [41], based on the whole dataset
2. Sort residus $r_n = -\log f(y_n, b_n; \hat{\theta}^{(p-1)}) = -\log P(Y_n = y_n, B_n, \hat{\theta}^{(p-1)})$ for $n = 1, \ldots, N$ with:

$$\begin{aligned} P(Y_n = y_n, B_n, \hat{\theta}^{(p-1)}) &= \textstyle\sum_{\omega_k} P(Y_n = y_n, B_n, \\ &\qquad X_n = \omega_k, \hat{\theta}^{(p-1)}) \\ &= \textstyle\sum_{\omega_k} P(X_n = \omega_k) b_n(k) \\ &\qquad f_k(y_n; \hat{\theta}^{(p-1)}) \end{aligned} \tag{15}$$

3. Define $H^{(p)} = \{y_{v(1)}, \ldots, y_{v(h)}\}$ the subset containing the h vectors with the lowest residus for $\hat{\theta}^{(p-1)}$
4. Compute $\hat{\theta}^{(p)} :=$ MLE using EM, based on $H^{(p)}$. We assign the likelihood of data considered an outliers to one, i.e., $f_k(y_n) = 1, \forall k$ in the HMC process. On the location where the data is considered an outlier, only prior distribution takes place in the labeling process. Calculation of the different probabilities becomes:

   - Forward probabilities:
     - $\alpha_1(k) = \pi_k f_k(y_1; \hat{\theta}^{(p)}) b_1(k)$
     - $\alpha_n(k) = \sum_{l=1}^{K} \alpha_{n-1}(l) a_{lk} f_k(y_n; \hat{\theta}^{(p)}) b_n(k)$ with $f_k(y_n, \hat{\theta}^{(p)}) = 1$ if y_n is considered as an outlier
   - Backward probabilities:
     - $\beta_N(k) = 1$
     - $\beta_n(k) = \sum_{l=1}^{K} \beta_{n+1}(l) a_{kl} f_l(y_{n+1}; \hat{\theta}^{(p)}) b_{n+1}(l)$ with $f_l(y_{n+1}, \hat{\theta}^{(p)}) = 1$ if y_{n+1} is considered as an outlier
   - A posteriori joint probabilities:

$$\begin{aligned} \xi_n(i, j) &= P(X_n = \omega_i, X_{n+1} = \omega_j | Y = y, B) \\ &= \frac{\alpha_{n-1}(j) a_{ji} f_i(y_n, \hat{\theta}^{(p)}) b_n(i) \beta_n(i)}{\sum_k \alpha_n(k)} \end{aligned}$$

   - A posteriori marginal probabilities:

$$\gamma_n(i) = P(X_1 = \omega_i | Y_1, \ldots, Y_N) = \frac{\alpha_n(i) \beta_n(i)}{\sum_j \alpha_N(j)}$$

   - $\mu_i = \frac{\sum_{n_1} \gamma_{n_1}(i) y_{n_1}}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$

- $\sigma_i = \frac{\sum_{n_1} \gamma_{n_1}(i)(y_{n_1}-\mu_i)(y_{n_1}-\mu_i)^t}{\sum_{n_1} \gamma_{n_1}(i)}$ with y_{n_1} belonging to the subset $H^{(p)}$

5. Back to step 2 until convergence of $H^{(p)}$

The main drawback of this approach is that the trimming parameter h representing the percentage of voxels used to estimate the parameters has to be fixed by the user. To carry out this problem, we propose to use an adaptative trimming parameter and a threshold s for the probability $P(Y_n = y_n, B_n, \theta)$ (15). At each iteration, the voxels for which the probability $P(Y_n = y_n, B_n, \theta)$ is lower than the threshold s (in practice $s = 0.02$) are considered outliers to the model and not included in HMC parameter estimation. In this case, the trimming parameter h will change at each iteration.

Outlier voxels also occur outside MS lesions, especially in the CSF class. Thus, to remove these outliers which are not MS lesions, a postprocessing step was added to our algorithm. Outliers for which the prior probability of CSF given by the atlas is higher than 0.5 were removed and lesions with a small volume ($3\,\text{mm}^3$) were excluded.

Results

Results Obtained for Brain Tissue Segmentation

In this section, we present some results obtained on real images. The 18 normal MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/. Only, the "hard" segmentation provided by experts is available in the Internet Brain Segmentation Repository (IBSR). The dimensions for the images were $256 \times 256 \times 128$. We use the Kappa index KI to evaluate the results:

$$\text{KI} = 2\frac{\#(\text{seg} \cap \text{GT})}{\#\text{seg} + \#\text{GT}} \tag{16}$$

where GT stands for ground truth map, seg for segmented map. Furthermore, $\#X$ represents the cardinal of X, $X \cap Y$ is the intersection between the sets X and Y. The better are the results, the closer to 100% are the criteria. To use this criterium, we need a segmentation into three classes. Therefore, after applying the proposed algorithm, we perform a hard segmentation, by assigning each voxel to the most representative class.

To evaluate the performance of this algorithm (HMC), we compare the proposed method with two other classic methods: SPM5[1] (statistical parametric mapping) [5], EMS[2] (expectation-maximization segmentation) [43]. SPM5 and EMS are two free

[1] http://www.fil.ion.ucl.ac.uk/spm.

[2] http://www.medicalimagecomputing.com/EMS/.

available reference software for brain MRI segmentation. Ashburner and Friston include tissue classification, bias field correction and image registration with tissue probability maps in the same model [5]. Van Leemput et al. propose a framework including bias field correction and tissue classification in an MRF model [43].

Each case of the IBSR database was segmented with the three different methods: SPM5, EMS, and our proposed method. An example of the results obtained on IBSR data is presented in Fig. 7. Figure 8 compares the performance of the different methods on each image of the IBSR dataset.

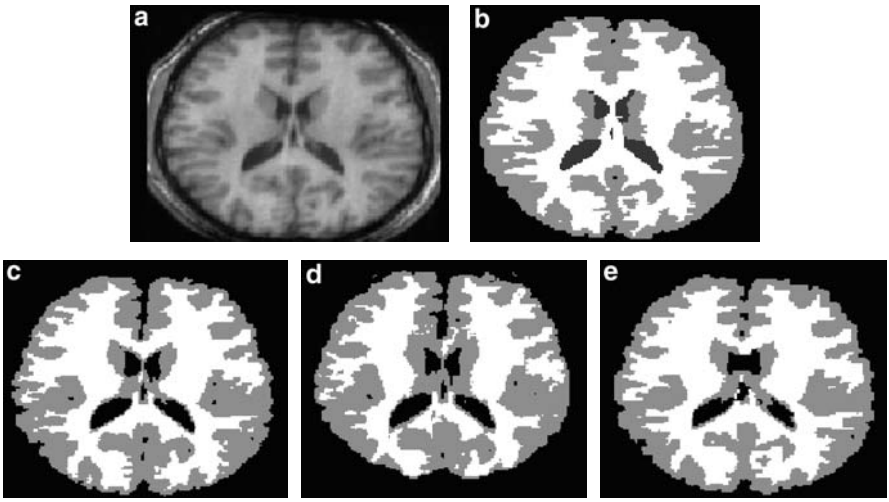


Fig. 7 Results obtained on the image IBSR 02. (**a**) Data to be segmented and (**b**) expert segmentation. (**c**), (**d**) and (**e**) Results obtained with SPM5, EMS and the proposed method, respectively. Kappa index obtained for this image with the different methods is presented in Fig. 8

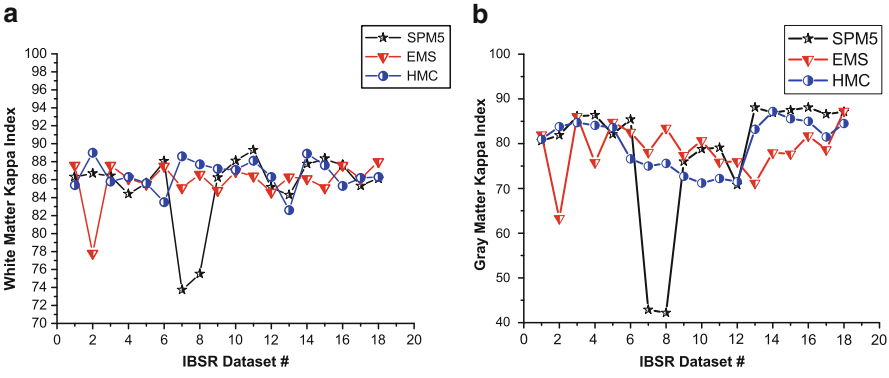


Fig. 8 Comparison of the results obtained on the IBSR dataset with the different methods: SPM5, EMS and the proposed method. Kappa index is reported for: (**a**) white matter, (**b**) gray matter

Results Obtained for MS Lesion Detection

In this section, we present some results obtained on real brain MR images for MS lesion detection. We applied the method on eight cases of the database using both T1 and Flair images. The results have been compared with two manual expert segmentations. To validate the method we use:

- The Kappa index KI
- The true positive fraction TPF
- The false positive fraction FPF

The Kappa indexes obtained for each case of the database are presented in Fig. 9 and the true and false positive fractions in Fig. 10. Concerning the Kappa index, the results we obtained are closer to those obtained by expert 2. The average Kappa index on the eight cases is 0.66 when compared with expert 2 and 0.62 when compared with expert 1. These results are slightly less than the average Kappa index comparing the segmentations of the two experts (0.74). Concerning the TPF and FPF, the results obtained by the method we proposed are closer to those obtained by expert 1. Expert 2 obtained an average TPF of 0.79 and an average FPF of 0.24 when

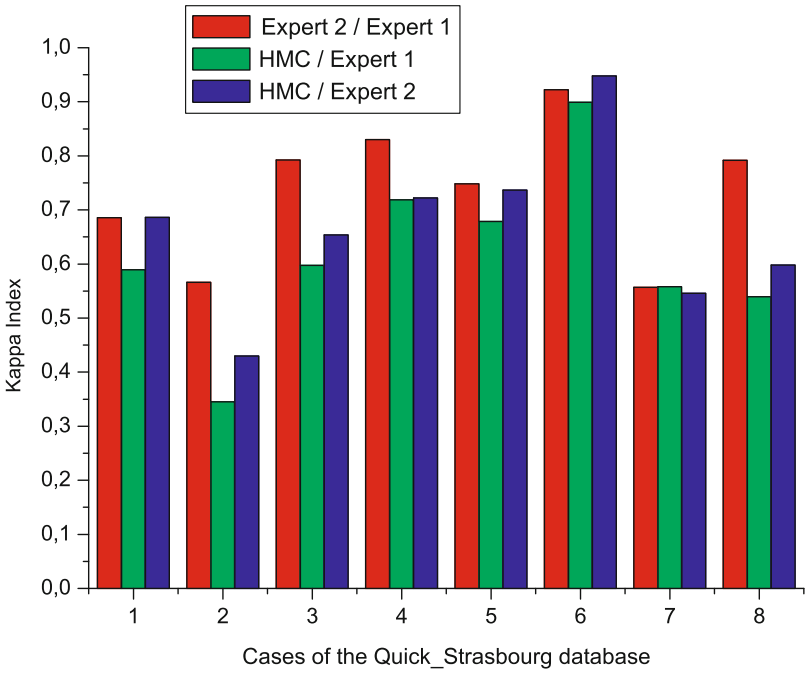


Fig. 9 Index Kappa obtained on each case of the database. We compared the results to each expert segmentation and the results of the experts together

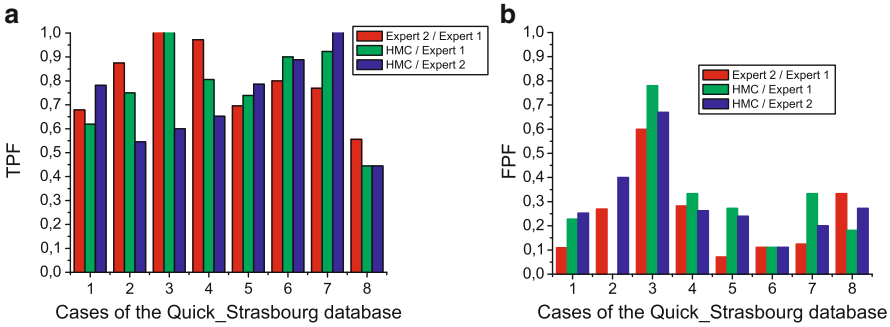


Fig. 10 Results obtained on each case of the database. (**a**) and (**b**) True positive fraction (TPF) and the false positive fraction (FPF), respectively

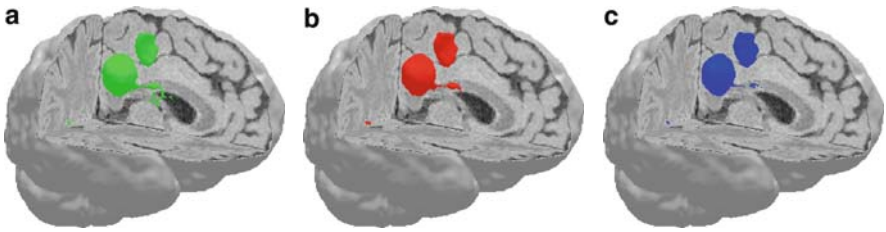


Fig. 11 3D representation of the results obtained of the case 06 of the database. (**a**), (**b**) and (**c**) Segmentations of expert 1, expert 2 and the HMC method, respectively

compared with expert 1. The proposed HMC method obtained an average TPF of 0.77 and an average FPF of 0.28 when compared with expert 1, whereas it obtained an average TPF of 0.71 and an average FPF of 0.3 when compared with expert 2.

Examples of the segmentations obtained on different cases of the database are presented in Figs. 11 and 12.

MS Lesion Refinement

In the previous section, we detected lesions as outliers to the model and a priori information brought by a probabilistic atlas was used. We now focus on the refinement on the contours of the detected lesions and thus, we will assume that the location of lesions has already been detected. We propose to extend the framework described in [15, 27] for 2D images segmentation, to 3D brain MRI segmentation. We first introduce the image model and the optimization scheme based on the minimum description length (MDL) principle. Section "Segmentation Algorithm" presents the proposed algorithm for 3D image segmentation. In section "Validation," results obtained on 3D brain MRI are shown.

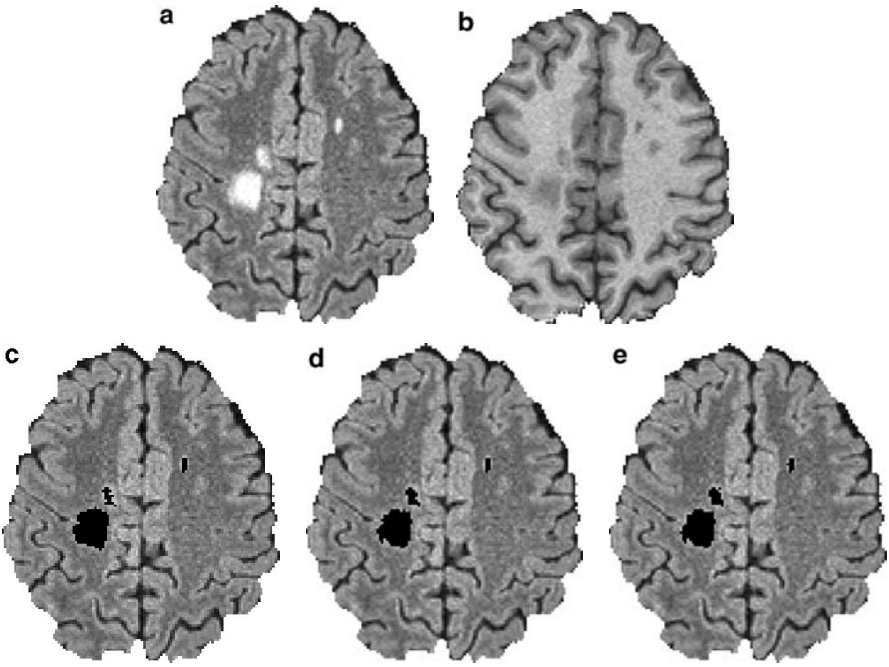


Fig. 12 Example of the results obtained of the case 06 of the database. (**a**) and (**b**) Flair and T1 images, respectively. (**c**), (**d**) and (**e**) Segmentations of expert 1, expert 2 and the HMC method, respectively

3D Statistical Active Contours

Image Model

Let us consider an image $\mathbf{s} = \{s(x, y, z)|(x, y, z) \in N_x \times N_y \times N_z\}$ composed of $N = N_x \times N_y \times N_z$ voxels, where $s(x, y, z)$ represents the gray level of voxel (x, y, z). We assume that this image is composed of two regions Ω_a and Ω_b corresponding respectively to the target object and to the background. Each region $\Omega_{l,l\in\{a,b\}}$ is supposed to be homogenous, i.e., all the voxels of the same region Ω_l are assumed to follow the same distribution P_{θ_l} where θ_l is the parameter vector of the probability density function (pdf) in region Ω_l. In this case, we assume that the gray levels in each region Ω_l follow a Gaussian distribution. $\theta_l = (\mu_l, \sigma_l)$ where μ_l and σ_l are respectively the mean and the standard deviation in Ω_l. $\theta = (\theta_a, \theta_b)$ denotes the two parameter vectors.

The image partitioning can be described by using the binary window function w: $w = \{w(x, y, z)|(x, y, z) \in N_x \times N_y \times N_z\}$ with $w(x, y, z) = 1$ within the shape and $w(x, y, z) = 0$ elsewhere.

Stochastic Complexity

The aim of segmentation consists in estimating the shape $\mathbf{w}$ of the target in the scene. The surface of the shape is defined by a k-triangular mesh modeling the contour between the target and the background. To estimate the number of triangles k defining the shape, we use the minimum description length (MDL) principle introduced by Rissanen [32]. This is based on the minimization of the number of bits Δ needed to describe the image. As the image we considered is composed of three parts (the background, the target and the contour), Δ is the sum of the three following terms: the length Δ_a required for the description of the target gray levels, the length Δ_b for the description of the background gray levels, and the length Δ_{w^k} needed for the description of the surface mesh composed of k triangles.

Thus, the stochastic complexity is composed of two terms: one corresponding to the code length associated to the image model and the other corresponding to the code length associated to the data coding knowing the model. Let us consider $\Delta(\mathbf{s}, w, \theta)$ the stochastic complexity needed to describe the image $\mathbf{s}$ associated to the partitioning ω and to the two pdfs defined by θ. We have:

$$\Delta(\mathbf{s}, w, \theta) = \Delta_G(w) + \Delta_L(\mathbf{s}|\theta, w) \tag{17}$$

with

- $\Delta_G(w)$ the number of nats needed to code the partitioning w

$$\Delta_G(w) = \Delta_{w^k} \tag{18}$$

- $\Delta_L(\mathbf{s}|\theta, w)$ the image coding term knowing the image, i.e., the coding term of the voxel gray levels in each region Ω_a and Ω_b.

$$\Delta_L(\mathbf{s}|\theta, w) = \Delta_a + \Delta_b \tag{19}$$

We will now detail the computation of each term.

Description Length of the Surface Mesh

Let us provide an approximation of the minimal code length needed to describe a partitioning w. The partitioning w is described using a triangular mesh composed of k triangles. To code one triangle, we need to encode a starting vertex and the segments of the triangle. As the image contains N voxels, the location of a vertex can be coded by $\log N$ nats. Thus, to code the mesh, we need to encode k starting vertices and the p segment vectors, where p denotes the total number of segments. The k starting vertices can be coded by $k \log N$ nats.

An approximation of the encoding of the segment vector coordinates can be obtained by using a Shannon code [36]. To determine this code, the probability distribution associated to the vectors is required. Let $d_u(i)$ be the u coordinates of the

segment number i: $d_u(i) = u_2(i) - u_1(i)$, where $u = x$ or y or z; $(x_1(i), y_1(i), z_1(i))$ and $(x_2(i), y_2(i), z_2(i))$ are the coordinates of the extremities of the segment i. Let us assume that $d_u(i)$ is distributed with a pdf $P_{m_u}(d_u)$. If the parameter m_u of the pdf $P_{m_u}(d_u)$ is known, the average code length needed to encode the components of the segment vectors can be approximated by the opposite of the log-likelihood $\mathcal{L}_e[\chi_u|m_u]$ where $\chi_u = (d_u(i))_{i\in[1,p]}$. Thus, we obtain the following code length for the mesh:

$$\Delta_G(w) = k\log N - \mathcal{L}_e[\chi_x|m_x] - \mathcal{L}_e[\chi_y|m_y] - \mathcal{L}_e[\chi_z|m_z] + \Delta(m_x) + \Delta(m_y) + \Delta(m_z) \tag{20}$$

where $\chi_x = (d_x(i))_{i\in[1,p]}$, $\chi_y = (d_y(i))_{i\in[1,p]}$, $\chi_z = (d_z(i))_{i\in[1,p]}$, and where $\Delta(m_x)$, $\Delta(m_y)$ et $\Delta(m_z)$ are the number of nats needed to encode the parameters m_x, m_y and m_z. According to [32], as $m_{u,u\in\{x,y,z\}}$ is estimated on a sampling of size p, it can be encoded with $\log\sqrt{p}$ nats. This leads to:

$$\Delta_G(w) = k\log N - \mathcal{L}_e[\chi_x|\hat{m}_x] - \mathcal{L}_e[\chi_y|\hat{m}_y] - \mathcal{L}_e[\chi_z|\hat{m}_z] + (3/2)\times\log p \tag{21}$$

To determine the pdf $P_{m_{u,u\in\{x,y,z\}}}(d_u)$ for the components of the segment vectors, the maximum entropy principle is applied [20]. Moreover, the statistical mean value m_u is assumed to be known. Thus, we have to find the pdf P_{m_u} maximizing the entropy $S = -\int_{-\infty}^{+\infty} P_{m_u}(x)\log P_{m_u}(x)\mathrm{d}x$ under the constraint that the absolute value of its mean equals to $m = \int_{-\infty}^{+\infty} |x|\,P_{m_u}(x)\mathrm{d}x$. The corresponding pdf is an exponential law, thus we obtain:

$$P_{m_u}(d_u) = \frac{1}{2m_u} e^{\frac{-|d_u|}{m_u}} \tag{22}$$

Then, we have:

$$\mathcal{L}_e[\chi|m] = \sum_{i=1}^{p} \log[P_m(d(i))] \tag{23}$$

If we replace m by its maximum likelihood (ML) estimate $\hat{m} = (1/p)\sum_{i=1}^{p} d(i)$, we obtain:

$$\mathcal{L}_e[\chi|m] = -p\log(2\hat{m}) - p \tag{24}$$

The code length to encode the mesh can then be deduced from (21):

$$\Delta_G(w) = k\log N + (3/2)\times\log p + p(3 + \log(2\hat{m}_x) + \log(2\hat{m}_y) + \log(2\hat{m}_z)) \tag{25}$$

Description Length of Voxel Gray Levels $\Delta_L(\mathbf{s}|\theta, w)$

Let us provide an approximation of Δ_a and Δ_b the description length needed to describe the voxel gray levels in each region Ω_a and Ω_b. Let us first consider the code length $\Delta_l(\theta_l)$ required to code the voxel gray levels in the region $\Omega_{l,l\in\{a,b\}}$

knowing the parameter vector θ_l. Shannon has shown that the average number of bits needed to describe a set x of random events corresponds to the entropy of the probability density function (pdf) [36]. Thus, the average number of bits $\Delta_{l,l\in\{a,b\}}$ needed to describe N_l random variables distributed with pdf $P_{\Theta_l}(x)$ is equal to $\Delta_l = N_l S_l$ where the entropy S_l is given by $S_l = -\int P_{\Theta_l}(x)\log_2[P_{\Theta_l}(x)]\mathrm{d}x - \log_2(q)$, q stands for the quantization step. As the contribution of q lies only in adding a constant term to the description length, we will not take it into account in the following description. The average code length (in nats) for an entropic encoding of the voxels' gray level in region Ω_l is given by [36]:

$$\Delta_l = -\sum_{(x,y,z)\in\Omega_l} \log P_{\Theta_l}(s(x,y,z)) \tag{26}$$

where $l = a$ or b. Thus, $\Delta_a + \Delta_b$ is the opposite of the log-likelihood of the image $\mathcal{L}_e[\mathbf{s}|w,\theta_a,\theta_b]$.

$$\Delta_L(\mathbf{s}|\theta,w) = -\mathcal{L}_e[\mathbf{s}|w,\theta_a,\theta_b] = -\mathcal{L}_e[\Omega_a|\theta_a] - \mathcal{L}_e[\Omega_b|\theta_b] \tag{27}$$

with $\mathcal{L}_e[\Omega_l|\theta_l]$ the opposite of the log-likelihood in the region Ω_l (where $l \in \{a,b\}$) :

$$\mathcal{L}_e[\Omega_l|\theta_l] = \sum_{(x,y,z)\in\Omega_l} \log P_{\theta_l}(s(x,y,z)) \tag{28}$$

$\mathcal{L}_e[\Omega_l|\theta_l]$ depends on the parameter vector θ_l of the pdf. As θ_l is unknown, it must be estimated. Let us consider $\hat{\theta}_l$ the maximum likelihood estimate of θ_l. By replacing θ_l by $\hat{\theta}_l$, we obtain the pseudo log-likelihood $\mathcal{L}_e[\Omega_l|\hat{\theta}_l]$ and equation (27) becomes:

$$\Delta_L(\mathbf{s}|\theta,w) = -\mathcal{L}_e[\Omega_a|\hat{\theta}_a] - \mathcal{L}_e[\Omega_b|\hat{\theta}_b] \tag{29}$$

We assume that the voxel gray levels in each region Ω_l follow a Gaussian pdf:

$$P_{\theta_l}(x) = \frac{1}{\sqrt{2\pi}\sigma_l}\exp^{-\frac{1}{2}\left(\frac{x_l-\mu_l}{\sigma_l}\right)^2} \tag{30}$$

where μ_l and σ_l are respectively the mean and the standard deviation in region Ω_l. Thus, the log-likelihood in region Ω_l can be written as:

$$\mathcal{L}_e[\Omega_l|\theta_l] = \sum_{(x,y,z)\in\Omega_l} \log P_{\theta_l}(s(x,y,z)) \tag{31}$$

$$= -\frac{N_l}{2}\log(2\pi) - N_l\log\sigma_l - \frac{1}{2\sigma_l^2}\sum_{(x,y,z)\in\Omega_l}(s(x,y,z)-\mu_l)^2 \tag{32}$$

with N_l the number of voxels in region Ω_l.

By replacing θ_l by $\hat{\theta}_l$, we obtain the pseudo log-likelihood $\mathcal{L}_e[\Omega_l|\hat{\theta}_l]$:

$$\mathcal{L}_e[\Omega_l|\hat{\theta}_l] = -\frac{N_l}{2}\log(2\pi) - N_l \log\hat{\sigma}_l - \frac{1}{2}N_l \tag{33}$$

$\hat{\mu_l}$ and $\hat{\sigma}_l^2$ are the empirical mean and the variance computed over the region Ω_l:

$$\hat{\mu_l} = \frac{1}{N_l}\sum_{(x,y,z)\in\Omega_l} s(x,y,z) \tag{34}$$

$$\hat{\sigma}_l^2 = \frac{1}{N_l}\sum_{(x,y,z)\in\Omega_l} (s(x,y,z) - \hat{\mu_l})^2 \tag{35}$$

Thus the gray level coding term can be written as:

$$\Delta_L(\mathbf{s}|\hat{\theta}, w) = \frac{1}{2}N_a \log\hat{\sigma}_a^2 + \frac{1}{2}N_b \log\hat{\sigma}_b^2 + K \tag{36}$$

with $N_{l,l\in\{a,b\}}$ the number of voxels in region $\Omega_{l,l\in\{a,b\}}$ and $\hat{\sigma}^2_{l,l\in\{a,b\}}$ the empirical variance in region $\Omega_{l,l\in\{a,b\}}$ and K a constant term.

Minimization of the Stochastic Complexity

Using the two terms $\Delta_G(w)$ and $\Delta_L(\mathbf{s}|\hat{\theta}, w)$ from (25) and (36), we can deduce the stochastic complexity $\Delta(\mathbf{s}, w, \hat{\theta})$ associated to a given partitioning w:

$$\begin{aligned}\Delta(\mathbf{s}, w, \hat{\theta}) = {} & k\log N + (3/2)\times\log p + p(3 + \log(2\hat{m}_x) + \log(2\hat{m}_y) + \log(2\hat{m}_z)) \\ & + \frac{1}{2}N_a \log\hat{\sigma}_a^2 + \frac{1}{2}N_b \log\hat{\sigma}_b^2\end{aligned} \tag{37}$$

In the following, we will denote the stochastic complexity $\Delta(w)$ to simplify the notation. The aim of segmentation consists in finding the function w^{MCS} minimizing the stochastic complexity $\Delta(w)$:

$$w^{\text{MCS}} = \arg\min_{w} \Delta(w) \tag{38}$$

Segmentation Algorithm

Optimization Scheme

Segmentation consists in finding the function $\mathbf{w}$ minimizing the stochastic complexity criterium $\Delta(w)$. To solve such an optimization problem, we use a multiresolution strategy based on two steps. First step lies in segmenting the object

by increasing the number of triangles and vertices to increase the 3D resolution and to obtain a mesh with an overestimated number of triangles k_0. Second step consists in reducing the complexity of the mesh obtained at the end of the first step. For this aim, each edge of the mesh is considered successively and we calculate the value of the criterion Δ' if this edge is removed. The edge that leads to the minimal value of Δ' is removed.

The mesh is initialized by a few number of triangles (for example like the cube presented in Fig. 16). To be less sensitive to the initialization, we adopt a two steps procedure. First step alternates vertices' moving and triangle subdivision to obtain a mesh with an overestimated number of triangles. During the subdivision of triangles, the value of the criterion Δ increases. The second step alternates nodes' moving and edge collapse to decrease the value of the criterion Δ. Triangles' subdivision, vertices' moving and edge collapse are presented in the next subsections.

Triangles' Subdivision

At the beginning of the segmentation algorithm, the mesh contains few triangles. Thus, the number of triangles is increased during the subdivision step. Each triangle is divided into four triangles by using the modified Butterfly scheme [48] to smooth the mesh. The triangles' subdivision step is presented in Fig. 13.

Nodes' Moving

The moving step consists in moving the nodes of the mesh to minimize the stochastic complexity. Each node is considered successively and moved. This move is accepted if it has lowered the stochastic complexity, or canceled otherwise. In practice, the number of moves is reduced to limit the computing time. For each node, we consider 26 moves with a fixed amplitude a (Fig. 14). The aim consists in finding the position corresponding to the lowest value of the criterion among the 27 potential positions (the 26 tested moves and the initial position). Thus, the amplitude of the move has to be set beforehand. In [19] for the 2D case, Galland proposed a multiresolution strategy to define the parameters such that the user does not need to set them. We have chosen to apply the same method in the 3D case. First, for each node P_i of the mesh, the chosen amplitude a_i is equal to $m_i/2$ where m_i is the mean

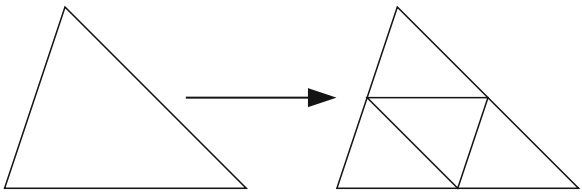


Fig. 13 Triangles' subdivision step

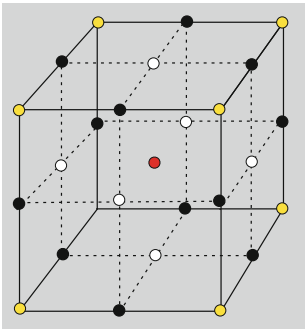


Fig. 14 The 26 potential moving and the initial position

length of the segments connected to the considered node P_i. Thus, for each node, the amplitude will be different. When any node can be moved, another moving step will be applied with all the amplitude a_i divided by 2. This process is carried out until each amplitude is less than a certain threshold $a_{\min}$, for example 1 voxel.

The main interest of this strategy consists in using a large amplitude at the beginning of the moving step to allow a fast convergence to the position of the contour without being blocked in a local minimum on one hand; and in decreasing this amplitude at the end of the convergence to estimate the position of the mesh with a better accuracy on the other hand.

The nodes' moving step is presented in Algorithm 3.

Algorithm 3 Nodes' moving step

1. For each node P_i, $a_i = m_i/2$
2. While one node has been moved, Do
   a. For each node P_i
      – For each of the 26 moves with amplitude a_i
        - Move the node $\rightarrow$ **w'**
        - Test if any crossing has been created
        - Compute the value of the criterion $\Delta(\mathbf{w'}, \mathbf{s})$ for this new mesh
      – Among these moves, keep the one with the lowest value of the criterion $\Delta(\mathbf{w'}_{\min}, \mathbf{s})$
      – Compare $\Delta(\mathbf{w'}_{\min}, \mathbf{s})$ and $\Delta(\mathbf{w}, \mathbf{s})$
        - If $\Delta(\mathbf{w'}_{\min}, \mathbf{s}) < \Delta(\mathbf{w}, \mathbf{s})$, apply the transformation: $\mathbf{w} \leftarrow \mathbf{w'}_{\min}$
        - Else cancel the move
3. If there is some nodes associated to an amplitude $a_i \geq 2a_{\min}$,
   a. For these nodes P_i, $a_i \leftarrow a_i/2$
   b. Go back in (2)
   Else the convergence is completed

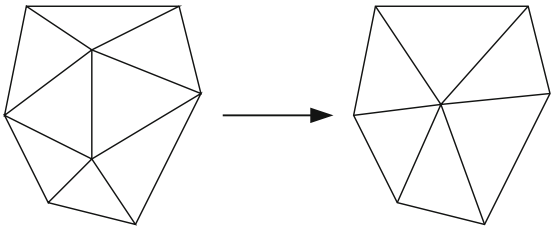


Fig. 15 Edge collapse

Edge Collapse

To simplify the mesh, we use an edge collapse transformation [24]. Each edge is considered successively and the edge which suppression leads to the lowest value of the criterion Δ is removed. This step leads to a distortion of the contour. Thus, after each edge collapse, we apply a nodes' moving step to adjust the mesh. This process is completed until the value of the criterion Δ does not decrease. The edge collapse step is presented in Fig. 15.

Segmentation Algorithm

The first step consists in alternating triangles' subdivision step and nodes' moving until the distance between two nodes is less than a minimal distance $d_{\min}$ (in practice $d_{\min} = 1$ voxel). Thus, we obtain a contour with an overestimated number of nodes and triangles. Then, a step which alternates edge collapse and nodes' moving is applied by minimizing the stochastic complexity to estimate the number of nodes of the mesh. The segmentation algorithm based on statistical active contours is presented in Algorithm 4.

Algorithm 4 Segmentation algorithm based on statistical active contours

1. $d_{\min} = 1$ voxel
2. Mesh initialization
3. While the distance between two nodes $d > d_{\min}$, Do
   - Triangles' subdivision (Fig. 13)
   - Nodes' moving (Algorithm 3)
4. While the stochastic complexity decreases, Do
   - Edge collapse (Fig. 15)
   - Nodes' moving (Algorithm 3)

Validation

Results on Synthetic Images

To validate the proposed method, tests have been first carried out on synthetic images. The 3D synthetic image is composed of a cone on a background corrupted by a Gaussian noise, parameters of which are described in Table 1. The initial mesh is a cube composed of 24 triangles and 14 vertices (Fig. 16b). Figure 16 shows the results obtained by the proposed method for the cone segmentation.

We tested the algorithm with three different positions of the initial mesh. We computed the error rates obtained in comparison with the ground truth by considering the voxels inside the contour. The error rates obtained are respectively 3.9, 4 and 4.1%.

Table 1 Parameters of the synthetic image

Class	Mean μ	Variance σ^2
Cone	1	0.1
Background	0	0.1

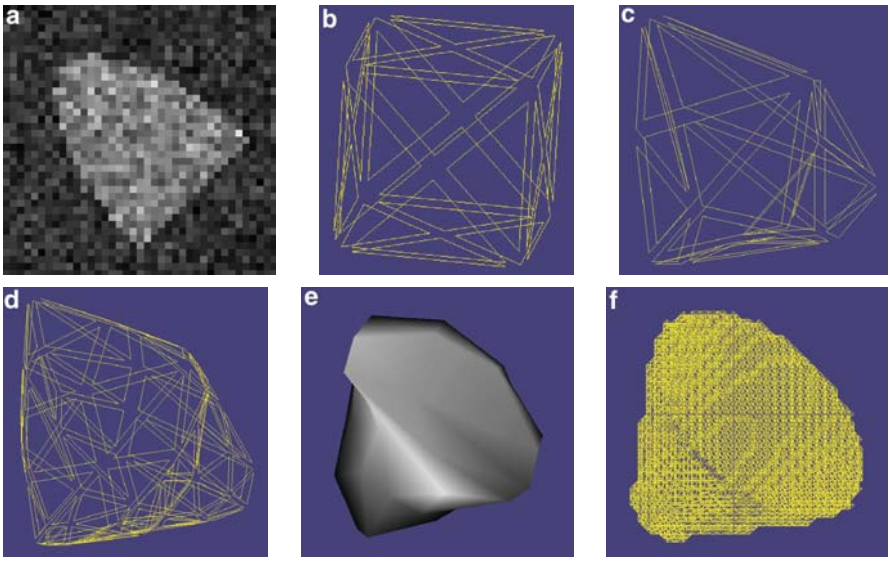


Fig. 16 Results obtained by the proposed method for cone segmentation. (**a**) Image to be segmented. (**b**) Initial mesh. (**c**) and (**d**) Results after 1 and 2 phases of nodes' moving and triangles subdivision, respectively. (**e**) Final result. (**f**) Ground truth

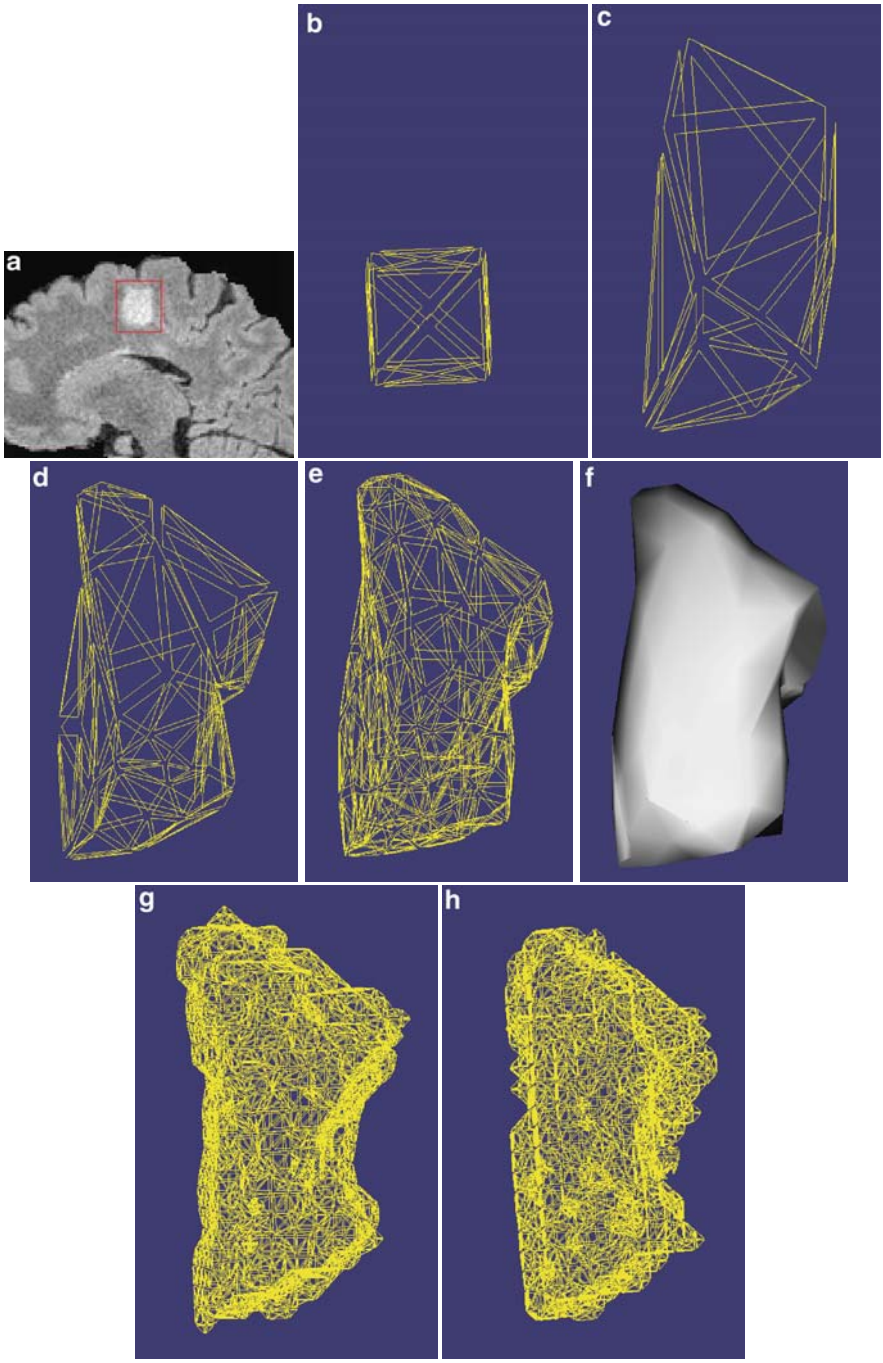


Fig. 17 Results obtained on a lesion. (**a**) Image we want to segment. (**b**) Initial mesh. (**c**), (**d**) and (**e**) Results after 1, 2 and 3 steps of nodes' moving and triangles subdivision, respectively. (**f**) Final mesh. (**g**) and (**h**) Manual segmentations of the experts

MS Lesion Segmentation

We present some results obtained on a MS lesion. The result we obtained is compared to the Marching cube representation of the manual segmentation of the experts [25]. The algorithm is initialized by a cube with 24 triangles and 14 vertices.

Results are presented in Fig. 17. We computed the error rates obtained compared to the manual segmentations of the experts. We obtained respectively 5.8 and 5.9% error rates.

Conclusion

We first proposed a new brain tissue segmentation method based on hidden Markov chains to include neighborhood information. This method takes into account both bias field inhomogeneities and partial volume effects by computing the proportion of each tissue in each voxel. This method also includes a priori information brought by a probabilistic atlas and enables to segment mono- and multimodal images. This method has been validated on both synthetic and real images and compared with other existing methods [12].

Then, this method was extended to MS lesion detection. The trimmed likelihood estimator (TLE) was used to estimate robustly the HMC model parameters and to detect the outliers. The use of a probabilistic atlas deriving from normal brains brings an information of the spatial localization of the outliers. To keep only the MS lesions, we applied constraints on these outliers. This method was tested on real images and compared with manual segmentation of two experts [10, 11].

We also search to obtain a more precise segmentation of the lesions by proposing a segmentation method based on 3D statistical active contours. This method is based on the principle of minimization of the stochastic complexity and consists in finding the function w minimizing the stochastic complexity Δ. We use a multiresolution strategy based on two steps to obtain the triangular mesh of the lesion.

Future work will consist in improving the initialization of the algorithm. The results of the MS lesion detection based on hidden Markov chains can be used to create the initial contour. This method could then be used to follow the evolution of the lesion in the time.

Acknowledgements We thank the Alsace Region and the CNRS for funding this research. We thank the Computational Geometry and Computer Graphics team of the LSIIT for their graphic modeling platform.

References

1. Aït-Ali L, Prima S, Hellier P, Carsin B, Edan G, Barillot C (2005) STREM: a robust multidimensional parametric method to segment MS lesions in MRI. In: Duncan J, Gerig G (eds) 8th international conference on medical image computing and computer-assisted intervention, MICCAI'2005, Springer, Palm Springs, Lecture Notes in Computer Science, vol 3749, pp 409–416
2. Al-Zubi S, Toennies K, Bodammer N, Hinrichs H (2002) Fusing Markov random fields with anatomical knowledge and shape based analysis to segment multiple sclerosis white matter lesions in magnetic resonance images of the brain. Bildverarbeitung für die Medizin pp 185–188
3. Ashburner J, Friston K (1997) Multimodal image coregistration and partitioning – a unified framework. NeuroImage 6(3):209–217
4. Ashburner J, Friston K (2000) Voxel-based morphometry – the methods. NeuroImage 11: 805–821
5. Ashburner J, Friston K (2005) Unified segmentation. NeuroImage 26:839–857
6. Bandoh Y, Kamata A (1999) An address generator for a 3-dimensional pseudo-Hilbert scan in a cuboid region. ICIP 496–500
7. Belaroussi B, Milles J, Carme S, Zhu YM, Benoit-Cattin H (2006) Intensity non-uniformity correction in MRI: existing method and their validation. Med Image Anal 10(2):234–246
8. Bosc M, Heitz F, Armspach JP (2003) Statistical atlas-based sub-voxel segmentation of 3D brain MRI. In: IEEE international conference on image processing (ICIP), Barcelone, Espagne, pp 1077–1080
9. Bovik A (2000) Handbook of image and video processing. Academic, San Diego
10. Bricq S, Collet C, Armspach JP (2008) Lesions detection on 3D brain MRI using trimmed likelihood estimator and probabilistic atlas. In: 5th IEEE international symposium on biomedical imaging ISBI'08, Paris, France
11. Bricq S, Collet C, Armspach JP (2008) Markovian segmentation of 3D brain MRI to detect multiple sclerosis lesions. In: IEEE international conference on image processing ICIP'08, San Diego, CA
12. Bricq S, Collet C, Armspach JP (2008) Unifying framework for multimodal brain MRI segmentation based on hidden Markov chains. Med Image Anal 12(6):639–652
13. Choi HS, Haynor DR, Kim YM (1991) Partial volume tissue classification of multichannel magnetic resonance images – a mixel model. IEEE Trans Med Images 10:395–407
14. Čížek P (2002) Robust estimation in nonlinear regression and limited dependent variable models. EconPapers
15. Delyon G, Galland F, Réfrégier P (2006) Minimal stochastic complexity image partitioning with unknown noise model. IEEE Trans Image Process 15(10):3207–3212
16. Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38
17. Dimova R, Neykov N (2004) Generalized D-fullness techniques for breakdown point study of the trimmed likelihood estimator with applications. In: Hubert M, Pison G, Struyf A, Van Aelst S (eds) Theory and applications of recent robust methods. Birkhauser, Basel, pp 83–91
18. Dugas-Phocion G, González Ballester MA, Malandain G, Lebrun C, Ayache N (2004) Improved EM-based tissue segmentation and partial volume effect quantification in multisequence brain MRI. In: Proceedings of MICCAI'04, Springer, Saint-Malo, France, Lecture Notes in Computer Science
19. Galland F (2004) Partition d'images par minimisation de la complexité stochastique et grille active: application à la segmentation d'images de radar à ouverture synthétique. PhD Thesis, Université d'Aix-Marseille III
20. Galland F, Bertaux N, Réfrégier P (2003) Minimum description length synthetic aperture radar image segmentation. IEEE Trans Image Process 12(9):995–1006

21. Gelman A, Carlin J, Stern H, Rubin D (2005) Bayesian data analysis. Chapman and Hall, New York
22. Guillemaud R, Brady M (1997) Estimating the bias field of MR images. IEEE Trans Med Imaging 16(3):238–251
23. Hadi A, Luceno A (1997) Maximum trimmed likelihood estimators: a unified approach, examples, and algorithms. Comput Stat Data Anal 25:251–272
24. Hoppe H (1996) Progressive meshes. In: ACM SIGGRAPH '96, pp 99–108
25. Lorensen WE, Cline HE (1987) Marching cubes: a high resolution 3D surface construction algorithm. Comput Graph 21(3):163–169
26. Marroquin J, Vemuri B, Botelo S et al (2002) An accurate and efficient Bayesian method for automatic segmentation of brain MRI. IEEE Trans Med Imaging 21(8):934–945
27. Martin P, Réfrégier P, Galland F, Guérault F (2006) Nonparametric statistical snake based on the minimum stochastic complexity. IEEE Trans Image Process 15(9):2762–2770
28. Neykov N, Müller C (2003) Breakdown point and computation of trimmed likelihood estimators in generalized linear models. In: Dutter R, Filzmoser P, Gatter U, Rousseeuw P (eds) Developments in robust statistics, Physica-Verlag, Heidelberg, pp 277–286
29. Neykov N, Neytchev P (1990) A robust alternative of the MLE. Compstat'90, pp 99–100
30. Noblet V, Heinrich C, Heitz F, Armspach JP (2005) 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. IEEE Trans Image Process 14(5):553–566
31. Pieczynski W (1992) Statistical image segmentation. Mach Graph Vis 1(2):261–268
32. Rissanen J (1989) Stochastic complexity in statistical inquiry. World Scientific, Singapore
33. Rousseeuw P, Leroy A (1987) Robust regression and outlier detection. Wiley, New York
34. Ruan S, Jaggi C, Fadili J, Bloyet D (2000) Brain tissue classification of magnetic resonance images using partial volume modeling. IEEE Trans Med Imaging 19(12):1179–1187
35. Ruan S, Moretti B, Fadili J, Bloyet D (2002) Fuzzy Markovian segmentation in application of magnetic resonance images. Comput Vis Image Underst 85:54–69
36. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
37. Shattuck DW, Sandor-Leahy SR, Schaper KA, Rottenberg DA, Leahy RM (2001) Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13:856–876
38. Sled JG, Zijdenbos AP (1998) A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17(1):87–97
39. Smith S (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155
40. Styner M, Brechbuhler C, Szekely G, Gerig G (2000) Parametric estimate of intensity inhomogeneities applied to MRI. IEEE Trans Med Imaging 19(3):153–165
41. Tanner M (1993) Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions. Springer, Berlin
42. Tohka J, Zijdenbos A, Evans A (2004) Fast and robust parameter estimation for statistical partial volume models in brain MRI. NeuroImage 23(1):84–97
43. Van Leemput K, Maes F, Vandermeulen D, Suetens P (1999) Automated model-based bias field correction of MR images of the brain. IEEE Trans Med Imaging 18(10):885–896
44. Van Leemput K, Maes F, Vandermeulen D, Suetens P (1999) Automated model-based tissue classification of MR images of the brain. IEEE Trans Med Imaging 18(10):897–908
45. Van Leemput K, Maes F, Vandermeulen D, Suetens P (2003) A unifying framework for partial volume segmentation of brain MR images. IEEE Trans Med Imaging 22(1):10–113
46. Vandev D, Neykov N (1993) Robust maximum likelihood in the Gaussian case. In: New Directions in Data Analysis and Robustness, Birkhäuser Verlag Basel, Switzerland, pp 259–264
47. Wells WM, Grimson WEL, Kikinis R, Jolesz FA (1996) Adaptative segmentation of MRI data. IEEE Trans Med Imaging 15(4):429–442
48. Zorin D, Schröder P, Sweldens W (1996) Interpolating subdivision for meshes with arbitrary topology. In: SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, ACM, New York, pp 189–192

Knowledge-Driven Recognition and Segmentation of Internal Brain Structures in 3D MRI

Isabelle Bloch

Introduction

The complexity of the human body and of its understanding through medical imaging techniques requires the development of image processing and interpretation methods that cope with this complexity. Recently, significant advances both in image acquisition techniques and in image processing have been performed. In particular, advanced and sophisticated image processing methods find in medical imaging a privileged field of applications. The major objectives concern help to diagnosis, therapy planning, surgical planning, patient's follow-up, morphometry, variability assessment, modeling, support for neuroscience applications, etc. This covers both clinical and research applications. Toward these aims, methods have to be developed to improve image quality, to perform segmentation and recognition of organs, pathologies, etc., to provide quantitative measures, to fuse multimodal image data, to provide numerical models, and to improve 3D visualization.

In this chapter, we summarize some of the work of our group on brain image interpretation, focusing on knowledge representation and its use to guide recognition and segmentation of internal brain structures in 3D MRI data. Instead of using shape models [1–3] or digital anatomical atlases [4–6], we rely on an explicit representation of structural information expressed as spatial relations between brain structures.

Since information and knowledge are subject to imprecision and vagueness, our mathematical models are based on fuzzy sets theory, as explained in section "Dealing with Imprecision Using Fuzzy Sets." In section "Knowledge Representation," we present our approach for knowledge representation. Fusion aspects are involved in several approaches we developed, and are summarized in section "Fusion." Finally we present the main lines of the recognition methods in section "Segmentation and Recognition," with a few illustrations.

I. Bloch (✉)
Telecom ParisTech, CNRS UMR 5141 LTCI, Paris, France
e-mail: isabelle.bloch@enst.fr

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_4,

Dealing with Imprecision Using Fuzzy Sets

Imprecision is often inherent to images, and its causes can be found at several levels: observed phenomenon (imprecise limits between structures or objects), acquisition process (limited resolution, numerical reconstruction methods), image processing steps (imprecision induced by a filtering for instance). Fuzzy sets have several advantages for representing such imprecision. First, they are able to represent several types of imprecision in images, as for instance imprecision in spatial location of objects, or imprecision in membership of an object to a class. For instance, partial volume effect, which occurs frequently in medical imaging, finds a consistent representation in fuzzy sets (membership degrees of a voxel to tissues or classes directly represent partial membership to the different tissues mixed up in this voxel, leading to a consistent modeling with respect to reality). Second, image information can be represented at different levels with fuzzy sets (local, regional, or global), as well as under different forms (numerical, or symbolic). For instance, classification based only on grey levels involves very local information (at the pixel level); introducing spatial coherence in the classification or relations between features involves regional information; and introducing relations between objects or regions for scene interpretation involves more global information and is related to the field of spatial reasoning. Third, the fuzzy set framework allows for the representation of very heterogeneous information, and is able to deal with information extracted directly from the images, as well as with information derived from some external knowledge, such as expert knowledge. This is exploited, in particular, in model-based pattern recognition, where fuzzy information extracted from the images is compared and matched to a model representing knowledge expressed in fuzzy terms.

Fuzzy set theory is of great interest to provide a consistent mathematical framework for all these aspects. It allows representing imprecision of objects, relations, knowledge, and aims at different levels of representation. It constitutes a unified framework for representing and processing both numerical and symbolic information, as well as structural information (constituted mainly by spatial relations in image processing). Therefore this theory can achieve tasks at several levels, from low level (e.g., grey-level-based classification) to high level (e.g., model-based structural recognition and scene interpretation). It provides a flexible framework for information fusion as well as powerful tools for reasoning and decision making.

Here fuzzy sets are mainly used for knowledge representation and for fusion.

Knowledge Representation

Models constitute an important source of information that provides generic knowledge, complementary to the actual patient's data and images. On one hand, biological, anatomical, or biomechanical models can be used to guide image interpretation. On the other hand, medical images can be exploited in order to build models of the human body, from an anatomical or functional point of view. Such models have been developed for specific organs, or for some functions such as breathing.

Let us here consider, more specifically, the first aspect. The main information contained in the images consists of properties of the objects and of relations between objects, both being used for pattern recognition and scene interpretation purposes. Relations between objects are particularly important since they carry structural information about the scene, by specifying the spatial arrangements between objects. These relations highly support structural recognition based on models. These models can be of iconic type, as an anatomical atlas, or of symbolic type, as linguistic descriptions, or ontologies. Although the use of iconic representations for normal structure recognition is well acknowledged, they remain difficult to exploit in pathological cases. Anatomical knowledge is also available in textbooks or dedicated Web sites, and is expressed mainly in linguistic form [7]. These models involve concepts that correspond to anatomical objects, their characteristics, or the spatial relations between them. Human experts use intensively such concepts and knowledge to recognize visually anatomical structures in images. This motivates their use in computer-aided image interpretation. Some attempts to formalize this knowledge have been recently performed, in particular in the form of ontologies (e.g., the foundational model of anatomy (FMA) [8]).

In our work, we concentrate mainly on spatial relations, which are strongly involved in linguistic descriptions. They constitute a very important information to guide the recognition of structures embedded in a complex environment, and are more stable and less prone to variability (even in pathological cases) than object characteristics such as shape or size. We proposed mathematical models of several spatial relations (adjacency, distances, directional relations, symmetry, between...) [9–14], in the framework of fuzzy sets theory, which proved useful to recognize thoracic and brain structures [15–17]. These models rely on fuzzy mathematical morphology operations [18], using fuzzy structuring elements representing the semantics of the spatial relations. For instance, dilating a (possibly fuzzy) reference object using such a structuring element provides a fuzzy region of space where the membership of each point expresses the degree to which the corresponding relation is satisfied at this point. Detailing the mathematical models is outside the scope of this chapter. These fuzzy representations can enrich anatomical ontologies and contribute to fill the semantic gap between symbolic concepts, as expressed in the ontology, and visual percepts, as extracted from the images. These ideas were used, in particular, in our segmentation and recognition methods [19,20]: a concept of the ontology is used for guiding the recognition by expressing its semantics as a fuzzy set, for instance, in the image domain or in an attribute domain, which can therefore be directly linked to image information.

Figure 1 [20] illustrates the ontological model. A part of the FMA concerning brain structures is extracted, and enriched with our ontology of spatial relations, establishing links between anatomical concepts and structural ones.

Figure 2 [21] illustrates the links between the different levels of the model and the reasoning. At the most abstract level, concepts are represented in ontologies and anatomical knowledge databases. An auxiliary graph structure helps linking this level with the spatial level, in which all concepts are translated into (fuzzy) regions. Finally the graph also guides the recognition in the images.

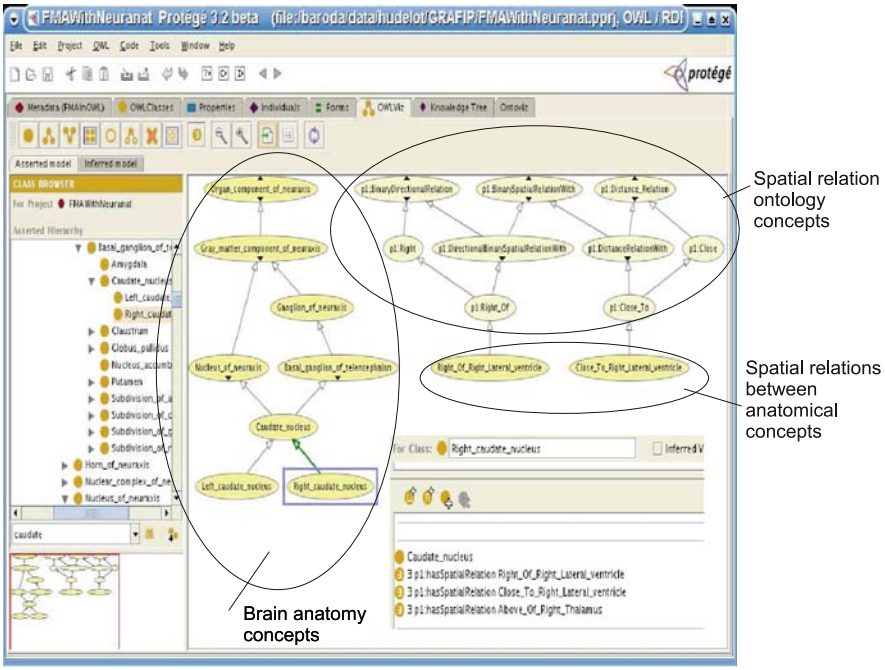


Fig. 1 Enriched ontology [20]

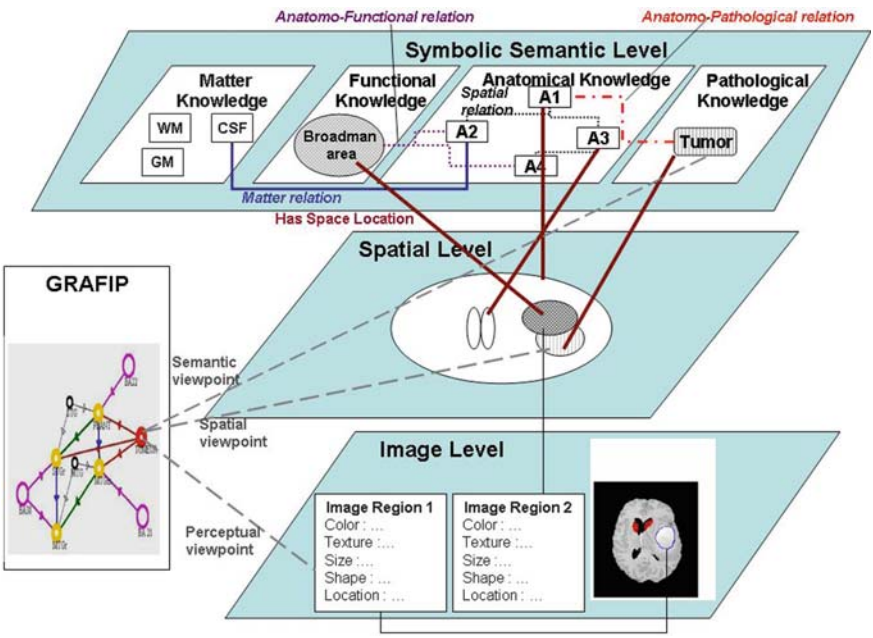


Fig. 2 Hierarchical organization [21]

Fusion

Information fusion becomes increasingly important in medical imaging due to the multiplication of imaging techniques. The information to be combined can be issued from several images (like multiecho MR images for instance), or from one image only, using, for instance, combination of several relations between objects or several features of the objects, or from images and a model, like an anatomical atlas, or knowledge expressed in linguistic form or as ontologies. This means that fusion can be involved at different levels of representations, from low level to higher structural one.

The advantages of fuzzy sets and possibilities rely in the variety of combination operators, offering a lot of flexibility in their choice, and which may deal with heterogeneous information [22, 23]. We proposed a classification of these operators with respect to their behavior (in terms of conjunctive, disjunctive, compromise [22]), the possible control of this behavior, their properties and their decisiveness, which proved to be useful for several applications in image processing [24]. It is of particular interest to note that, unlike other data fusion theories (like Bayesian or Dempster–Shafer combination), fuzzy sets provide a great flexibility in the choice of the operator, that can be adapted to any situation at hand. The combination process can be done at several levels of information representation, from pixel level to higher level. A noticeable advantage of this approach is that it is able to combine heterogeneous information, like it is usually the case in multi-image fusion.

At a numerical level, the typical application is multisource classification. We present an example of image fusion problem in brain imaging, where dual-echo brain MR images are combined in order to provide a classification of the brain into three classes: brain, ventricles and CSF, and pathology. These images are shown in Fig. 3. The membership functions for these classes have been estimated in as unsupervised way on both images and then used in a fuzzy fusion

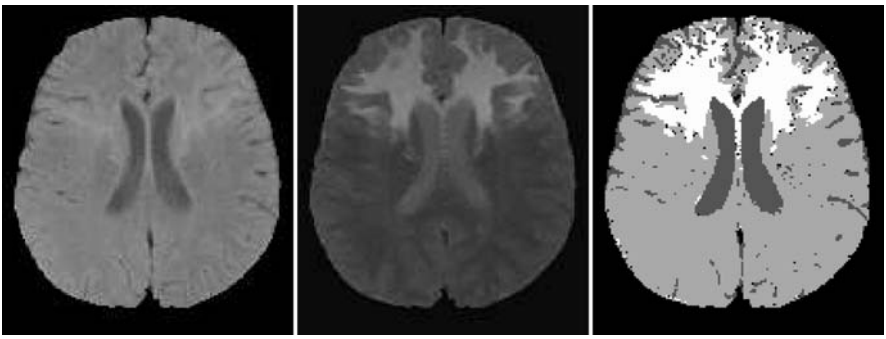


Fig. 3 Dual echo MR image of the brain, showing three main classes: brain, ventricles, and pathology (the *white area* on the middle image). *Right*: final decision after fuzzy combination (note that the decision is taken at each pixel individually, without spatial regularization)

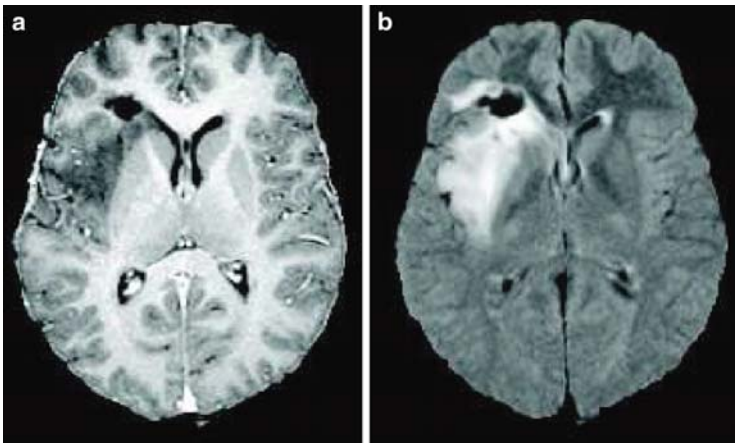


Fig. 4 T1-weighted (**a**) and T2-weighted (**b**) images of a brain with tumor and edema. While the pathological area, defined as the tumor and the edema, appears as one single region, very contrasted from the other tissues in the T2-weighted image (*in white*), the same area decomposes in the T1-weighted image into the tumor (*dark grey*) and the surrounding edema (*lighter grey*, hardly distinguishable from grey matter)

scheme [25]. Fusion operators are adapted to the image content for each class. After the combination, the decision is made according to the maximum of membership values. The result is shown in Fig. 3 (right).

Another example concerns the fusion of heterogeneity measures in the framework of multiphase level sets methods. We proposed to use fuzzy operators in this framework for extending the approach of [26, 27] to the multichannel case [28, 29]. This approach has been applied to the segmentation of brain tissues in pathological T1 and T2 MRI images obtained using SPGR and FLAIR sequences (Fig. 4). An example showing the contours of the tumors is illustrated in Fig. 11.

At a structural level, the operations defined on fuzzy objects as well as the relations between fuzzy objects can serve as a basis for structural recognition [15–17]. Fusion can also be involved in methods based on constraint satisfaction problem (CSP) approaches, as in [30]. Examples will be provided in the next section.

A noticeable advantage of fuzzy fusion is that it is able to combine heterogeneous information, like is the case when dealing with higher level approaches, where several types of knowledge and information with different semantics have to be combined, and to avoid to define a more or less arbitrary and questionable metric between pieces of information.

Let us give a few examples. If we have different constraints about an object (for instance concerning the relations it should have with respect to another object) which have all to be satisfied, these constraints can be combined using a t-norm (a conjunction). If one object has to satisfy one relation or another one then a disjunction represented by a t-conorm has to be used. This occurs for instance when

two symmetrical structures with respect to the reference object can be found (this situation often occurs in medical imaging). Mean operators can be used to combine several estimations and try to find a compromise between them. Associative symmetrical sums can be used for reinforcing the dynamics between high and low membership degrees. Importance of a constraint or reliabilities can be easily introduced in adaptive operators, and so on.

Segmentation and Recognition

The methods we develop in our group for segmentation and recognition of 3D structures in medical images can be seen as spatial reasoning processes. Two main components of this domain are spatial knowledge representation and reasoning. In particular spatial relations constitute an important part of the knowledge we have to handle, as explained before. Imprecision is often attached to spatial reasoning in images, and can occur at different levels, from knowledge to the type of question we want to answer. The reasoning component includes fusion of heterogeneous spatial knowledge, decision making, inference, recognition. Two types of questions are raised when dealing with spatial relations:

1. Given two objects (possibly fuzzy), assess the degree to which a relation is satisfied
2. Given one reference object, define the area of space in which a relation to this reference is satisfied (to some degree). This is illustrated in Fig. 5.

In order to answer these questions and address both representation and reasoning issues, we rely on three different frameworks and their combination:

- Mathematical morphology, which is an algebraic theory that has extensions to fuzzy sets and to logical formulas, and can elegantly unify the representation of several types of relations

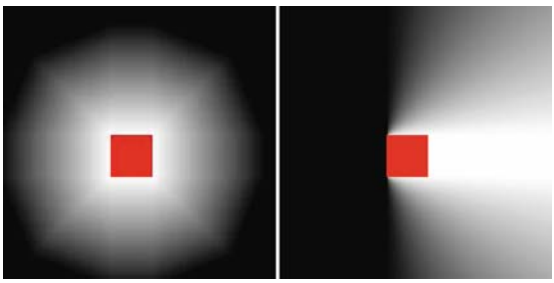


Fig. 5 Spatial representations of the relation "close to the central square" (*left*) and "right to the square" (*right*). Grey levels represent the degree of satisfaction of the relation

- Fuzzy set theory, which has powerful features to represent imprecision at different levels, to combine heterogeneous information and to make decisions
- Formal logics and the attached reasoning and inference power

The association of these three frameworks for spatial reasoning is an original contribution of our work [31].

Three types of approaches have been developed, based on the two types of questions.

Global Approach

In the first approach, which relies on the first type of question, spatial relations evaluated between spatial entities (typically objects or regions) are considered as attributes in a graph. The model is a graph derived from an anatomical atlas. Each node represents an anatomical structure, and edges represent spatial relations between these structures. A data graph is constructed from the MRI image where recognition has to be performed. Each node represents a region obtained from a segmentation method. Since it is difficult to segment directly the objects, usually the graph is based on an oversegmentation of the image, for instance, based on watersheds. Attributes are computed as for the model. The use of fuzzy relations is particularly useful in order to be less sensitive to the segmentation step.

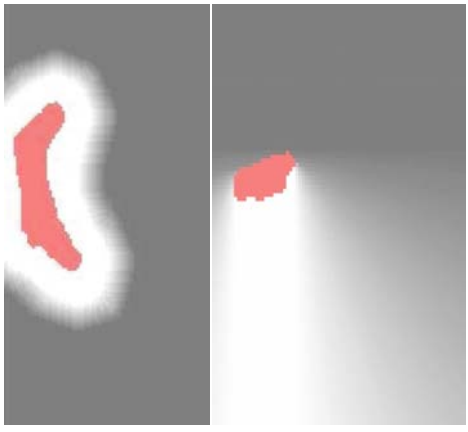
One important problem to be solved then is graph matching. Because of the schematic aspect of the model and the difficulty to segment the image into meaningful entities, no isomorphism can be expected between both graphs. In particular, several regions of the image can be assigned to the same node of the model graph. Such problems call for inexact graph matching. In general, it consists in finding a morphism, which furthermore optimizes an objective function based on similarities between attributes. Here the fusion applies not directly on the relations but on the similarities between them (see section "Fusion"). A weighted mean operator allows giving more importance to the edges, which show less variability between subjects and therefore constitute stronger anchors for guiding recognition. The morphism aims at preserving the graph structure, while the objective function privileges the association between nodes, respectively, between edges, with similar attribute values. This approach can benefit from the huge literature on fuzzy comparison tools (see, e.g., [32]) and from recent developments on fuzzy morphisms [33]. The optimization is not an easy task since the problem is NP-hard. Genetic algorithms, estimation of distribution algorithms, and tree search methods have been developed toward this aim [34–36].

Sequential Approach

In the second type of approach, relying on the second type of question, we use spatial representations of spatial knowledge [16, 17]. It consists in first recognizing simple structures (typically brain and lateral ventricles), and then progressively more and more difficult structures, based on relations between these structures and previously recognized ones. The order in which structures can be recognized can be provided by the user, or estimated as suggested in [37, 38] by computing an optimal path in a graph, based on a suitable metrics incorporating prior knowledge and image information such as saliency. Each relation describing the structure to be recognized is translated into a spatial fuzzy set representing the area satisfying this relation, to some degree (see Figs. 5 and 6).

The fuzzy sets representing all relations involved in the recognition process are combined using a numerical fusion operator. While we first used an atlas in [16], this constraint has been relaxed in our recent work [17]. This presents two main advantages: the high computation time associated with the computation of a deformation field between the atlas and the image is left aside and the procedure is potentially more robust because it uses only knowledge expressed in symbolic form, which is generic instead of being built from a single individual as in the iconic atlas.

Finally, a refinement stage is introduced using a deformable model. This stage uses an initial classification (using a low level approach based on grey levels) as a starting point and has the potential to correct possible imperfections of the previous stage together with regularizing the contours of structures. This deformable model makes use of a fusion of heterogeneous knowledge: edge information derived from the image, regularization constraints, and spatial relations contained in the linguistic description. All pieces of information are combined in the energy of a parametric or geodesic deformable model [17, 39]. For instance, the caudate nucleus can be recognized based on its grey level (roughly known depending on the type of acquisition), and, more importantly, on its relations to the lateral ventricles



close to the lateral ventricle below the lateral ventricle

Fig. 6 Fuzzy spatial representations of spatial relations

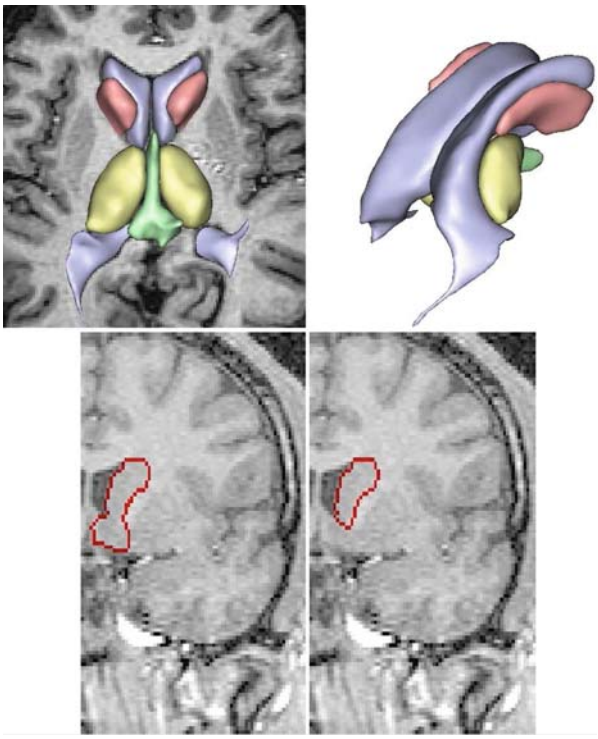


Fig. 7 Segmentation and recognition results obtained for the lateral ventricles, third ventricle, caudate nuclei and thalami by integrating spatial relations in 3D deformable models [17]. Illustration of the importance of spatial relations in the deformable model: in the case of caudate nucleus, the force derived from spatial relations prevents the model to grow below the lower limit of the structure (*left*: result obtained without this force, *right*: with this force)

(exterior and close to them). Here, the primary role of spatial relations is to prevent the deformable model from progressing beyond the limit of structures with weak boundaries.

Figure 7 shows 3D views of some cerebral objects recognized in an MR image with this method. In particular, the importance of spatial relations is illustrated in the case of the caudate nucleus. The lower part of this structure has a very weakly defined boundary and the use of a spatial relation is essential to achieve a good segmentation [17].

One of the advantages of this approach is that it can be extended to pathological cases, since spatial relations remain quite stable in the presence of pathologies, unlike shapes and absolute locations. Moreover, it is possible to learn the parameters of the relations, and their stability according to the type of pathology [19, 40]. Two examples of segmentation and recognition results in pathological cases are shown in Fig. 8, based on a segmentation of the tumor (based on fuzzy classification) [41].

Figure 9 summarizes the main steps of this approach.

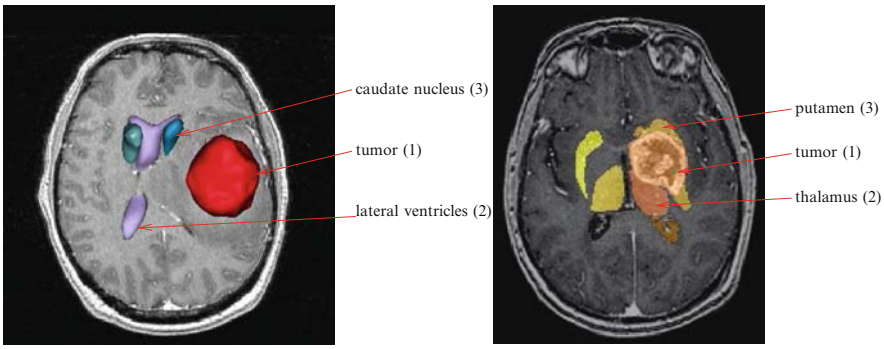


Fig. 8 Examples of segmentation and recognition in pathological cases [19, 40]

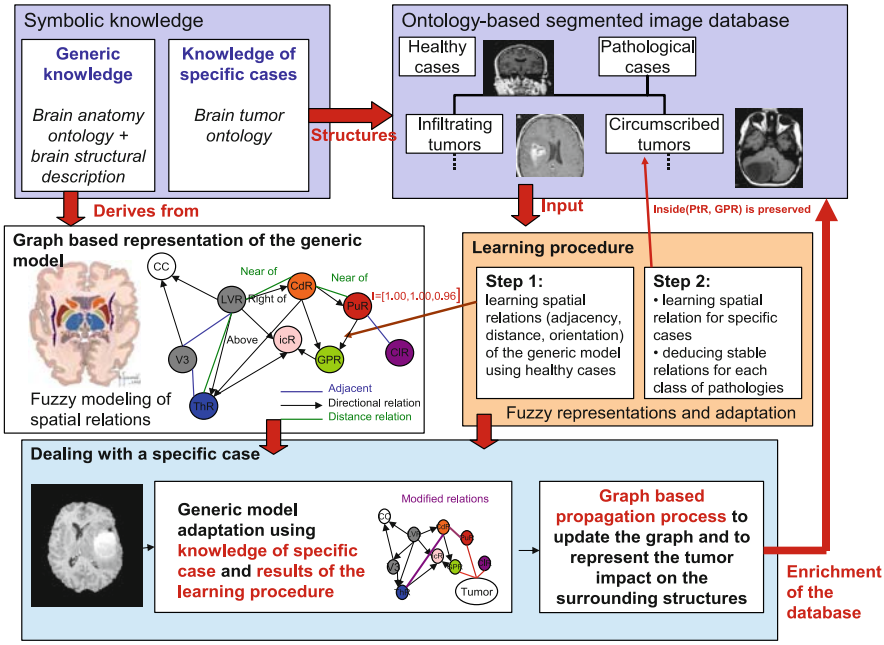


Fig. 9 Method overview [19, 20]

CSP Based Approach

An approach, that can be considered as intermediate between the two previous ones, was recently proposed in [30, 42]. The structural model is still a graph representing the anatomical structures and their spatial relations, represented by fuzzy models. A constraint network is built from the relations and the problem is then expressed as a CSP. The proposed solver is an original global method extracting a solution (i.e., the assignment of a region of space to an anatomical structure) according to the

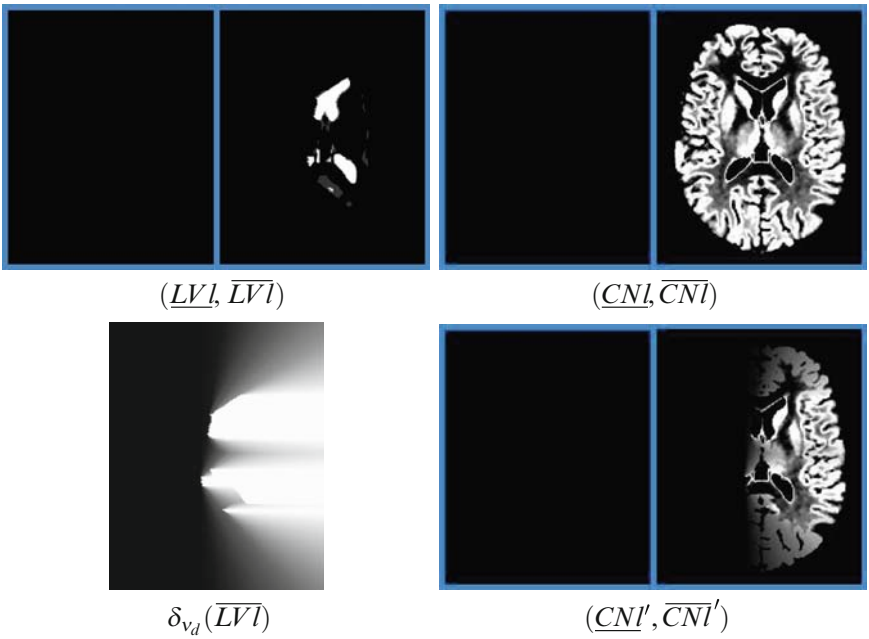


Fig. 10 Illustration on one slice of the bounds of the left lateral ventricle and the left caudate nucleus. Once the relation "the caudate nucleus is to the left of the ventricles" is taken into account using a specific propagator, the bounds of the caudate nucleus are updated, thus reducing the possible regions for this structure [30, 42]

relations of the structural model. We proposed to progressively reduce the solution domain by excluding assignments that are inconsistent with a constraint network derived from the structural model (see Fig. 10). The final segmentation of each structure is again performed using a deformable model. An example is provided in Fig. 12.

Outlook

In this chapter, several examples illustrating the potential of knowledge-based and fuzzy methods for brain imaging have been described. The association of a mathematical framework for modeling imprecision at different levels and of artificial intelligence methods for representing concepts and knowledge and for reasoning on them seems to be a very interesting current trend, where promising results are expected in a near future.

Our current work consists in developing further the methods for pathological cases, including quantification and patient's follow-up. Other applications such as

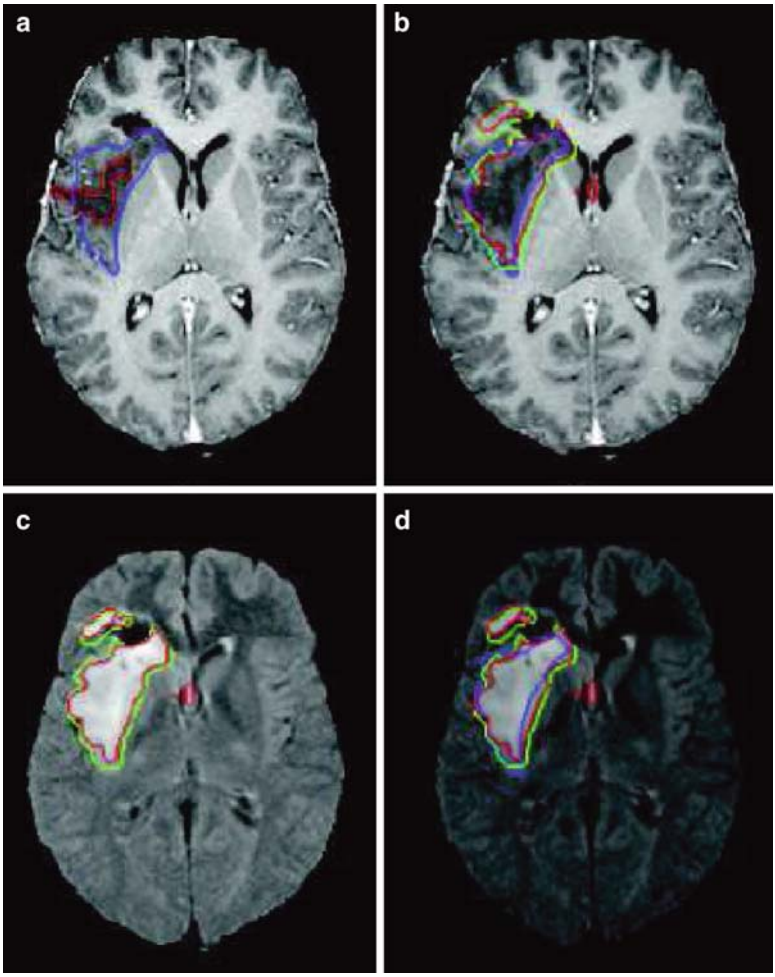


Fig. 11 (**a**) Contour of the tumor in single-channel SPGR segmentation in red, compared to the manual SPGR contour in *blue*; (**b**) tumor in multichannel segmentation in red, compared to the manual SPGR contour in *blue* and the manual FLAIR contour in *green*, registered on SPGR image; (**c**) tumor in FLAIR segmentation in *red*, compared to the manual FLAIR contour in *green*; (**d**) tumor in multichannel segmentation in *red*, compared to the manual SPGR contour in *blue* and the manual FLAIR contour in *green*, registered on FLAIR image [29]

heart imaging, high-resolution vascular imaging, fetal imaging are also under study. On a more methodological point of view, more complex spatial relations are still being developed, including their learning from image databases, as well as spatial reasoning schemes.

Our methodology starts from linguistic descriptions of anatomical knowledge and then goes on with mathematical modeling, associated algorithms, and

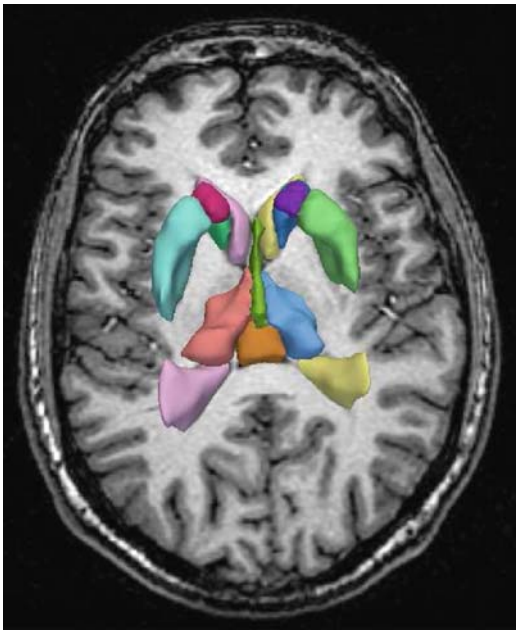


Fig. 12 3D view of recognition results for the caudate nuclei, putamen, ventricles, and thalami, superimposed on one slice of the original MRI data [30, 42]

applications. Conversely, segmentation and recognition results could lead to automatic descriptions of the image content in linguistic terms. This is also part of our ongoing work.

Acknowledgements The work described in this chapter is mostly joint work with several colleagues, including Elsa Angelini, Jamal Atif, Bénédicte Batrancourt, Olivier Colliot [43], Geoffroy Fouquier, Thierry Géraud [44], Céline Hudelot, Vincent Israel-Jost, Hassan Khotanlou [45], Olivier Nempont [42], Aymeric Perchant [46], Nathalie Richard. It has been performed in collaboration with several hospitals in Paris (Pitié-Salpétriére, Sainte-Anne, Saint Vincent de Paul) and laboratories (LENA-CNRS, LIMSI-CNRS, ENST Bretagne, IFR 49, universities of Sao Paulo, Bogota, Columbia...). It benefited from funding from Région Ile-de-France, ParisTech, INCA, ANR, Institut TELECOM...

References

1. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23(6):681–685
2. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models-their training and application. Comput Vis Image Underst 61(1):38–59
3. Leventon ME, Grimson WEL, Faugeras O (2000) Statistical shape influence in geodesic active contours. In: IEEE conference on computer vision and pattern recognition, CVPR, vol 1. pp 316–323
4. Broit C (1981) Optimal registration of deformed images. PhD thesis, University of Pennsylvania, Philadelphia, PA

5. Dawant BM, Hartmann SL, Thirion JP, Maes F, Vandermeulen D, Demaerel P (1999) Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. I. Methodology and validation on normal subjects. IEEE Trans Med Imaging 18(10):909–916
6. Iosifescu DV, Shenton ME, Warfield SK, Kikinis R, Dengler J, Jolesz FA, Mc-Carley RW (1997) An automated registration algorithm for measuring MRI subcortical brain structures. Neuroimage 6(1):13–25
7. Waxman SG (2000) Correlative neuroanatomy, 24th edn. McGraw-Hill, New York
8. Rosse C, Mejino JLV (2003) A reference ontology for bioinformatics: The foundational model of anatomy. J Biomed Inf 36:478–500
9. Bloch I (1999) Fuzzy relative position between objects in image processing: a morphological approach. IEEE Trans Pattern Anal Mach Intell 21(7):657–664
10. Bloch I (1999) On Fuzzy distances and their use in image processing under imprecision. Pattern Recognit 32(11):1873–1895
11. Bloch I (2005) Fuzzy spatial relationships for image processing and interpretation: A review. Image Vis Comput 23(2):89–110
12. Bloch I, Colliot O, Cesar R (2006) On the ternary spatial relation between. IEEE Trans Syst Man Cybern SMC-B 36(2):312–327
13. Bloch I, Maître H, Anvari M (1997) Fuzzy adjacency between image objects. Int J Uncertain, Fuzziness Knowl-Based Syst 5(6):615–653
14. Colliot O, Tuzikov A, Cesar R, Bloch I (2004) Approximate reflectional symmetries of fuzzy objects with an application in model-based object recognition. Fuzzy Sets Syst 147:141–163
15. Bloch I, Colliot O, Camara O, Géraud T (2005) Fusion of spatial relationships for guiding recognition. example of brain structure recognition in 3D MRI. Pattern Recognit Lett 26: 449–457
16. Bloch I, Géraud T, Maître H (2003) Representation and fusion of heterogeneous fuzzy information in the 3d space for model-based structural recognition – application to 3D brain imaging. Artif Intell 148:141–175
17. Colliot O, Camara O, Bloch I (2006) Integration of Fuzzy spatial relations in deformable models – application to brain MRI segmentation. Pattern Recognit 39:1401–1414
18. Bloch I, Maître H (1995) Fuzzy mathematical morphologies: a comparative study. Pattern Recognit 28(9):1341–1387
19. Atif J, Hudelot C, Fouquier G, Bloch I, Angelini E. From generic knowledge to specific reasoning for medical image interpretation using graph-based representations. In: International joint conference on artificial intelligence IJCAI'07, Hyderabad, India, January 2007, pp 224–229
20. Hudelot C, Atif J, Bloch I (2008) Fuzzy spatial relation ontology for image interpretation. Fuzzy Sets Syst 159:1929–1951
21. Hudelot C, Atif J, Nempont O, Batrancourt B, Angelini E, Bloch I (2006) GRAFIP: a framework for the representation of healthy and pathological anatomical and functional cerebral information. In: Human brain mapping, Florence, Italy, June 2006
22. Dubois D, Prade H (1985) A review of fuzzy set aggregation connectives. Inf Sci 36:85–121
23. Yager RR (1991) Connectives and quantifiers in fuzzy sets. Fuzzy Sets Syst 40:39–75
24. Bloch I (1996) Information combination operators for data fusion: A comparative review with classification. IEEE Trans Syst Man Cybern 26(1):52–67
25. Bloch I, Aurdal L, Bijno D, Muller J (1997) Estimation of class membership functions for grey-level based image fusion. In: ICIP'97, vol III. Santa Barbara, CA, October 1997, pp 268–271
26. Sandberg B, Chan TF (2005) A logic framework for active contours on multi-channel images. J Vis Commun Image Represent 16:333–358
27. Vese LA, Chan TF (2002) A multiphase level set framework for image segmentation using the Mumford and Shah model. Int J Comput Vis 50:271–293
28. Israel-Jost V, Breton E, Angelini E, Choquet P, Bloch I, Constantinesco A (2008) Vectorial multi-phase mouse brain tumor segmentation in T1-T2 MRI. In: IEEE international symposium on biomedical imaging (ISBI), Paris, France, May 2008, pp 5–8

29. Israel-Jost V, Daron J, Angelini E, Bloch I. Multi-phase and multi-channel region segmentation and application in brain MRI. Technical report, UCLA, CAM08–75, November 2008
30. Nempont O, Atif J, Angelini E, Bloch I. Structure segmentation and recognition in images guided by structural constraint propagation. In: European conference on artificial intelligence ECAI, Patras, Greece, July 2008, pp 621–625
31. Bloch I (2006) Spatial reasoning under Imprecision using fuzzy set theory, formal logics and mathematical morphology. Int J Approx Reason 41:77–95
32. Bouchon-Meunier B, Rifqi M, Bothorel S (1996) Towards general measures of comparison of objects. Fuzzy Sets Syst84(2):143–153
33. Perchant A, Bloch I (2002) Fuzzy morphisms between graphs. Fuzzy Sets Syst 128(2):149–168
34. Bengoetxea E, Larranaga P, Bloch I, Perchant A, Boeres C (2002) Inexact graph matching by means of estimation of distribution algorithms. Pattern Recognit 35:2867–2880
35. Cesar R, Bengoetxea E, Bloch I (2002) Inexact graph matching using stochastic optimization techniques for facial feature recognition. In: International conference on pattern recognition ICPR 2002, vol 2. Quebec, August 2002, pp 465–468
36. Perchant A, Boeres C, Bloch I, Roux M, Ribeiro C (1999) Model-based scene recognition using graph fuzzy homomorphism solved by genetic algorithm. In: GbR'99 2nd international workshop on graph-based representations in pattern recognition, Castle of Haindorf, Austria, pp 61–70
37. Fouquier G, Atif J, Bloch I (2007) Local reasoning in Fuzzy attributes graphs for optimizing sequential segmentation. In 6th IAPR-TC15 workshop on graph-based representations in pattern recognition, GbR'07, vol LNCS 4538. Alicante, Spain, June 2007, pp 138–147
38. Fouquier G, Atif J, Bloch I (2008) Sequential spatial reasoning in images based on pre-attention mechanisms and fuzzy attribute graphs. In: European conference on artificial intelligence ECAI, Patras, Greece, July 2008, pp 611–615
39. Nempont O, Atif J, Angelini E, Bloch I. Combining radiometric and spatial structural information in a new metric for minimal surface segmentation. In: Information processing in medical imaging (IPMI 2007), vol LNCS 4584. Kerkrade, The Netherlands, July 2007, pp 283–295
40. Khotanlou H, Atif J, Angelini E, Duffau H, Bloch I. Adaptive segmentation of internal brain structures in pathological MR images depending on tumor types. In: IEEE international symposium on biomedical imaging (ISBI), Washington, DC, April 2007, pp 588–591
41. Khotanlou H, Colliot O, Atif J, Bloch I (2009) 3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models. Fuzzy Sets Syst 160:1457–1473
42. Nempont O Modèles structurels flous et propagation de contraintes pour la segmentation et la reconnaissance d'objets dans les images. Application aux structures nor-males et pathologiques du cerveau en IRM. PhD thesis, Ecole Nationale Supérieure des Télécommunications, March 2009
43. Colliot O (2003) Représentation, évaluation et utilisation de relations spatiales pour l'interprétation d'images. Application à la reconnaissance de structures anatomiques en imagerie mé dicale. PhD thesis, Ecole Nationale Supérieure des Télécommunications, ENST 2003E036, September 2003
44. Géraud T. Segmentation des structures internes du cerveau en imagerie par résonance magnétique tridimensionnelle. PhD thesis, École Nationale Supérieure des Télécommunications, ENST 98E012, June 1998
45. Khotanlou H. Segmentation 3D de tumeurs et de structures internes du cerveau en IRM. PhD thesis, École Nationale Supérieure des Télécommunications ENST2008E005, February 2008
46. Perchant A Morphisme de graphes d'attributs flous pour la reconnaissance structurelle de scènes. PhD thesis, École Nationale Supérieure des Télécommunications, ENST00E046, November 2000

New Dimensions in Diagnostic Imaging of the Aorta

Jean Bismuth, Christof Karmonik, Dipan Shah, Mark G. Davies, and Alan B. Lumsden

Conventional Imaging

Computed Tomography

First surgical applications of X-ray imaging can be traced back as early as to the Spanish-American war in 1898, where Elizabeth Fleischmann (San Francisco) was recognized by the surgeon general of the United States for using images (film plates) obtained under different exposure angles, to triangulate pieces of shrapnel in the body of her patients [1]. In the 1920s, the first patents were issued for imaging a single plane or slice of the human body, soon called tomographs (tomos (greek): section or cut and graphein: to write) and the first instrument was built in 1931 [1]. In 1928, Jean Kieffer, a French Immigrant, devised the design of an X-ray apparatus that would yield a sharp image of the region of interest with the surrounding areas blurred out. The first such device, named the Kieffer laminagaph, was not built until 1937. With the help of others it was developed at the Mallinckrodt Institute of Radiology at Washington University in St. Louis, [1].

With the advent of the minicomputers in the late 1970s, the modality of computed tomography was born. Pioneers of this area were Sir Godfrey Hounsfield (EMI, Central Research UK) and Allan McLeod Comack (Tufts University, USA). Both shared the Nobel Prize in Medicine (1979) for their discoveries. The original 1971 prototype took 160 readings in a range of 180°, where each reading (or scan) took a few minutes. Algebraic reconstruction algorithms then took 2.5 h to compute the final image [2].

Since these early days, medical X-ray imaging has shown tremendous progress. Commercial CT scanners now provide axial images of the thorax including the heart in 0.6 s even eliminating the need of breath-holding and a true temporal resolution

J. Bismuth (✉), C. Karmonik, D. Shah, M.G. Davies, and A.B. Lumsden
The Methodist DeBakey Heart & Vascular Center, Houston, TX 77030, USA
e-mail: JBismuth@tmhs.org

C. Karmonik
The Methodist Hospital Neurological Institute, Houston, TX 77030, USA

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_5,

of 75 ms can be used to freeze the heart motion for visualizing coronary arteries (http://health.siemens.com/ct_applications/healthier_ct/20_sub-msv-heart.html, accessed June 16th 2009).

Magnetic Resonance Imaging

Although the effect of nuclear magnetism has been known since the early 1900s, imaging applications were not realized until the early 1970s. Similar to CT, computers are necessary to calculate an image from the huge amount of data that is acquired during these imaging procedures. In contrast to CT, where tissue absorption of X-rays is exploited to obtain image contrast and where the position of the image is determined by the position of the X-ray source and the detector, magnetic resonance relies on radiowaves that are emitted by the protons inside the patient as a response to a prior radiowave transmission. Image contrast is achieved by the variation of the response due to the chemical environment (or tissue types) of the protons and the image location is obtained through a sophisticated manipulation of the radiowave transmission and receptions. Although the main components of MRI were developed in the 1970s, in 2003 Paul Lauterbur (University of Illinois at Urbana-Champaign, USA) and Sir Peter Mansfield (University of Nottingham, UK) were awarded the Nobel Prize in Physiology or Medicine for their contributions. In contrast to CT or CTA, MRI has the capability of imaging physiological processes in addition to anatomical structures. In addition, the radiowaves that are used do not rely on ionizing radiation such as CT and therefore are much better suited for repeated application (serial imaging) or for use in vulnerable populations such as children or pregnant women.

Aortic Dissections

An Aortic dissection is a tear in the wall of the aorta that causes blood to flow between the layers of the wall of the aorta and forces the layers apart [3]. Aortic dissection is a medical emergency and can quickly lead to death, even with optimal treatment. If the dissection tears the aorta completely open (through all three layers), massive and rapid blood loss occurs. Aortic dissections resulting in rupture have an 80% mortality rate and 50% of patients die before they even reach the hospital. Aortic dissection is thought to represent the most common form of immediately life-threatening aortic pathology or aortic catastrophe [4]). Complications of aortic dissection include the development of aortic aneurysms. If these reach 6 cm in width, surgery is indicated as there is a significant risk of rupture [5]. An TB-AD presenting less than 2 weeks from symptom onset is defined as acute, while those that have been present 2 weeks or longer are classified as chronic

[6, 7]. Dissections confined to the descending aorta (type B, TB-AD) have better in-hospital survival than those involving the ascending aorta (type A). Up to 89% of patients with uncomplicated type B dissections survive to hospital discharge after receiving effective antihypertensive therapy [3, 8]. However, despite a low in-hospital mortality rate, the short- and long-term prognosis of patients with type B acute aortic dissection after discharge from the hospital is heterogeneous, with reported survival rates ranging from 56 to 92% at 1 year and from 48 to 82% at 5 years [8]. Given the variable prognosis of type B acute aortic dissection under current management strategies, new predictors of poor outcomes are needed.

At onset, the intra-arterial septum (IS) is believed to be thin and highly mobile. At a later stage, the IS will have thickened, as elastin is replaced by collagen and consequently, mobility will have decreased. Aortic dissection is an uncommon but highly lethal disease, affecting young people with an incidence of approximately (can also be reported as 3.5 per 100,000 persons –5.2 for men and 2.2 for women) 2,000 cases per year in the United States and 3,000 per year in Europe [9]. Often there is no indication that an aortic dissection is imminent, and it can occur suddenly. If there are symptoms, they include back pain or difficulty swallowing (as the dissected aorta pushes on the spinal cord or the esophagus). If not immediately treated, the mortality rate for acute Type A aortic dissection increases by approximately 1% per hour over the first 48 h and may reach 70% at 1 week. As many as 90% of untreated patients, who suffer aortic dissection, die within 3 months of presentation [10]. The condition made world headlines in 2003 when actor John Ritter died of an acute aortic dissection. Often, aortic dissections are caused by aortic aneurysms, or abnormally widened and weakened sections of a blood vessel wall.

Type III/Stanford B aortic dissections (dissections confined to the descending aorta, TB-AD) are commonly considered less lethal than aortic dissection involving the ascending aorta at onset. Disturbingly, however, a recent study utilizing data from a contemporary registry of acute B-AD revealed that the long-term prognosis after hospital discharge of patients with type B acute aortic dissection is heterogeneous, with reported survival rates ranging from 56 to 92% at 1 year and from 48 to 82% at 5 years [3, 8].

Given the variable prognosis of type B acute aortic dissection under current management strategies, predictors of poor outcomes are urgently needed and an improved classification scheme of TB-AD based on imaging and leading to improved outcome is warranted.

Finally, we have only recently been able to offer an alternative approach to the management of TB-AD, TEVAR. This involves utilizing an endograft to exclude the entry tear of the dissection in the descending aorta. The thought is that by doing this one is able to alter the natural course of these dissections and prevent complications such as malperfusion syndrome (poor blood flow to the arteries supplying the intestines and kidneys) as well as the development of aneurysms. Current preliminary literature seems to support TEVAR, but the results vary widely [11–13].

MRI of Aortic Dissections

Contrast-Enhanced Imaging

To evaluate the geometry of TB-AD a contrast-enhanced acquisition (ce-MRI) is routinely acquired in one breath-hold using intravenous contrast. Typical imaging parameters based on our experience are: sagittal orientation, slice thickness 1.75 mm, matrix 384 × 193, in-plane resolution 1.04 mm × 1.04 mm. The image data is provided as consecutive slices (Fig. 1).

New MRI acquisition sequences now allow the acquisition of 3D image data every few seconds. This technique, also known as time-resolved MR angiographic imaging, combines the advantage of large anatomic coverage together with the visualization of flow advancement [14]. This technique is available at our institution in an implementation called time-resolved imaging with stochastic trajectories (TWIST) on a Siemens 1.5 T Avanto Magnetom magnetic resonance scanner. Examples of TWIST images acquired for the human aorta are shown in Fig. 2.

2D Phase Contrast Magnetic Resonance Imaging

A well-established technique to visualize motion with MRI is 2-D pcMRI. With this technique, at several (typically 12–20) time points during the cardiac cycle,

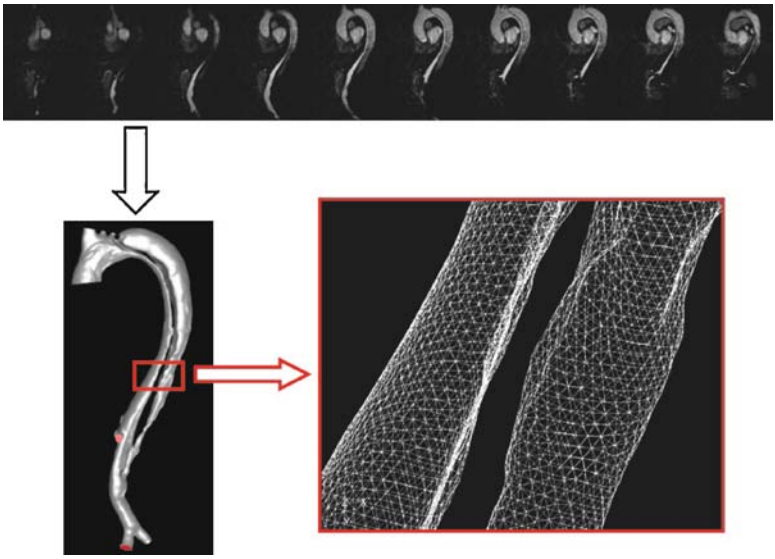


Fig. 1 *Top*: Selected slices of a 3D ce-MRI acqusition from a patient with TB-AD. *Bottom left*: 3D surface reconstruction depicting well the true and false lumen. *Bottom right*: Detail of computational mesh from red box denoted in image on *left*

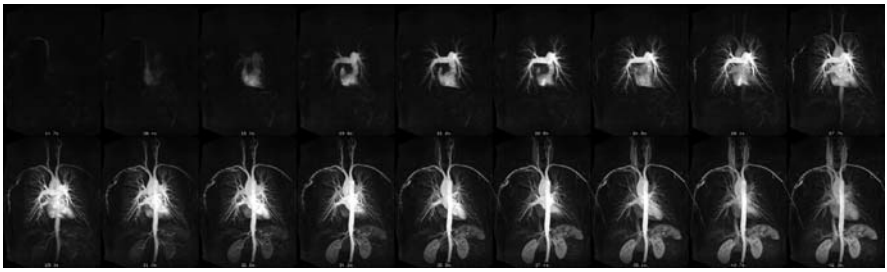


Fig. 2 Maximum intensity projections (MIP) of 18 3D image datasets acquired during contrast passage (administered intravenously in the left arm) every 1.6 s. Contrast first reaches the right ventricle, then the pulmonary arteries and veins, then the left ventricle and is ejected into the aorta. Contrast is then noted in the kidneys and spleen and finally in the inferior vena cava

images are obtained. These images are usually averaged over 10–15 cardiac cycles (one breath-hold). 2-D pcMRI has previously been shown to successfully measure changes in the aortic diameter for aortic dissections [15].

2-D pcMRI is a method that creates images where the signal intensity is derived from flow effects (or moving magnetization) by applying flow-encoding gradients as part of the MRI pulse sequence [16]. These gradients translate the velocity of the moving spins (e.g., the blood) into phase values of the magnetization which can then be visualized in the MRI phase images. To map the phase difference to a velocity, the operator has to prescribe an additional parameter, commonly called VENC (velocity encoding). If the VENC is too small, flow aliasing results may make the images unsuitable for flow analysis and if the VENC is too large, sensitivity is too poor, resulting in poor contrast to noise ratio and large errors in the flow analysis. For normal blood circulation, the maximum velocities in the major cerebral vessels are tabulated and these values can then be used to adjust the VENC. In pathologies such as stenoses, aneurysms, and arterio-vascular malformations, (AVM) however, maximum velocities can be increased several fold which makes it more difficult to select an optimum VENC value.

The 2-D pcMRI technique provides time-resolved magnitude images and phase difference images. The magnitude images can be used to study structural changes occurring during the cardiac cycle, such as IS motion. The phase difference images (acquired with the appropriate VENC value) allow for the quantification of aortic flow (provided the images were acquired perpendicular to the long axis of the aorta). Therefore, IS motion and aortic volumetric flow rates can therefore be measured with one acquisition.

Blood pressures cannot be measured with MRI. Pressure measurements in the true and false lumen of an aortic dissection have been successfully performed in our institution using a pressure sensor mounted on an intra-arterial catheter during EVAR treatment of B-AD. This procedure is invasive, however, and therefore prohibitive for use as a screening method. Computational simulation, which is justified if a parameter is not readily accessible via measurement, is therefore a better approach for obtaining information about luminal pressure differences in B-AD.

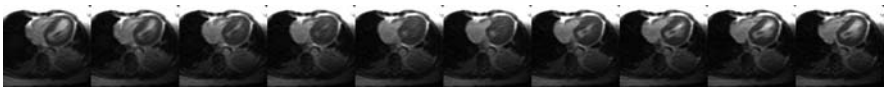


Fig. 3 Real-time MRI axial images (image data acquired every 59 ms) of a thrombosed TB-AD. The thrombus in the false lumen is visible as a hypo-intense crescent-shape area immediately posterior to the wall of the left ventricle

Real-Time Magnetic Resonance Imaging

While cardiac gated MRI techniques are well suited to display period motion, random, i.e., nonperiodic motion is not accessible by these methods as the random variation of the location of structures will lead to blurring in the resulting image. The design of rapid imaging sequences now makes it acquire an MRI image in 60–100 ms, where this value is determined by a trade-off of spatial resolution in the image. Figure 3 shows real-time MRI images in an axial orientation acquired from a patient diagnosed with TB-AD. Imaging revealed the presence of thrombus in the false lumen.

Computational Fluid Dynamics

Computational fluid dynamics (CFD) techniques have advanced from early studies with simplified geometries [17] to sophisticated approaches including patient-specific geometries and flows [18–21]. In particular, fluid structure interactions have gained interest to incorporate aortic wall motion into these simulations [20]. Image segmentation techniques and meshing algorithms have now advanced to a point where specific geometries of human arteries can be routinely extracted from MRI imaging data [18, 22, 23]. Patient-specific inflow waveforms derived from 2D pcMRI have been successfully used in CFD simulations [18, 23, 24]. The necessity for the use of measured patient-specific flow rates in the CFD simulations has been demonstrated as well: in a CFD study of an aneurysm of the anterior communicating artery, Karmonik et al. demonstrated that deviations of 43% in the wall shear stress magnitude in the aneurysm wall may result if the inflow into the aneurysm is not known [23].

While the technique of CFD for simulating laminar flow is well understood and has been constantly improved for many decades, validation of CFD results for blood flow in-vivo is still needed. A recent comparison study of velocities measured with 2D pcMRI and simulated with CFD in intracranial sidewall aneurysms of the internal carotid artery showed excellent agreement and is encouraging for future applications of CFD [25].

The application of CFD methods for patient-specific modeling consists of sophisticated postprocessing steps. In a first step, a 3D image data set is identified, that is suitable for providing the geometrical information to extract the aortic wall

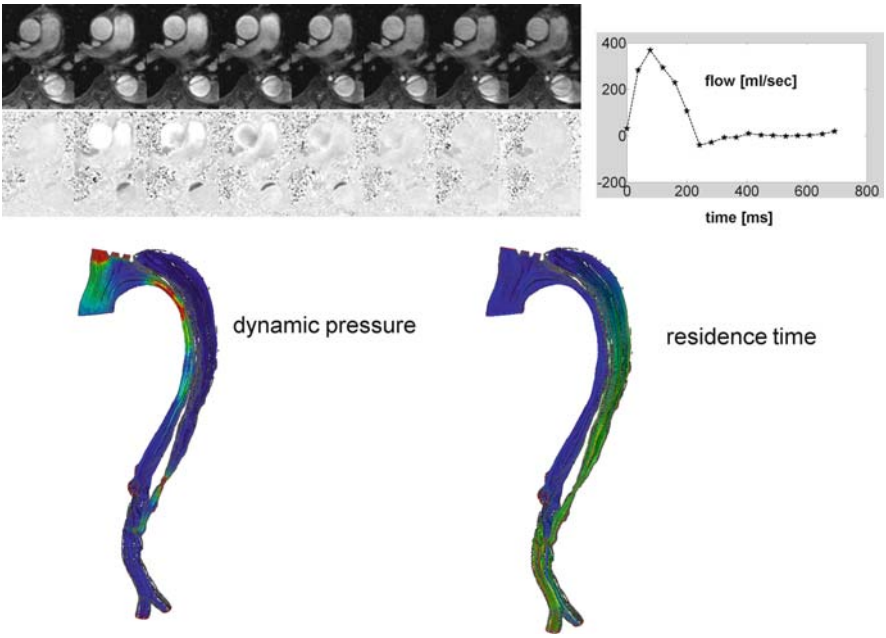


Fig. 4 *Top*: 2D pcMRI images (magnitude images showing anatomical structure on top, flow images showing velocity information below). revealing helical flow pattern in the ascending aorta. Aortic volumetric flow waveform is shown on right. *Bottom*: Pathlines of flow during time of maximum aortic flow colored by dynamic pressure (*blue*: low, *red*: high) on left and particle residence time on right (*blue*: low, *red*: high). The dynamic pressure reveals stronger flow in true lumen, the residence time shows flow delays in the false lumen, most probably due to stenosis immediately posterior to re-entrance tear

of the true and the false lumen (Fig. 1). A 3D surface reconstruction of the aorta and the main branching arteries is then constructed. This surface is then imported into commercial or self-written software that breaks up the lumen of the artery in small segments. In the example discussed here, the GAMBIT software program (Ansys, Inc.) was utilized to create 360,275 volume elements inside the aortic lumen. The volume of these elements ranged from 3.56×10^{-11} to 2.02×10^{-9} m^3. This computational mesh can then serve as input into a solver software for calculating flow patterns in the aortic lumen. To obtain realistic flow information, a physiologic realistic flow waveform has to be used as input into these simulations. 2D pcMRI can provide this waveform (Fig. 4).

Motion Analysis of Aortic Septum

At present, outcome of endovascular treatment (EVAR) of type III B aortic dissections (TB-AD) is difficult to predict. Recently, a technique has been presented

for quantifying intraseptal IS motion from cine 2D pcMRI images toward a new classification of TB-AD based on IS mobility [26]. A semiautomated algorithm was developed to parameterize true aortic lumen boundary (tAB) and IS motion based on cross-sectional images of the aorta. The maximum distension, maximum contraction, and average displacement of the tAB and the IS relative to the center of mass of the true aortic boundary points (ALC) were provided. Correlations of the average motion of the ALC, the tAB, and the IS with the aortic blood flow waveform were calculated. Figure 5 illustrates the main concept of this approach. Similar approaches have been reported: Minami et al. quantified diameter changes of the true and false lumen of aortic dissections potentially indicative of IS movement [15]. In a series of ten consecutive cases of acute aortic dissection (six Stanford type A and four Stanford type B) the movement of the intimal flap was characterized by a flap movement index (FMI). Maximum and minimum aortic diameters (ADmax, ADmin), as well as maximum and minimum false lumen diameters (FDmax, FDmin) were measured. FMI was defined as (FDmax/ADmax) – (FDmin/ADmin)]/(FDmin/ADmin) × 100 (%).The so-defined FMI ranged between 4 and 87% (mean 26.3+/−8.6%). The authors concluded that IS movement as characterized by the FMI is a potentially useful parameter in cases with acute aortic dissection to determine whether or not emergency surgery is indicated [15].

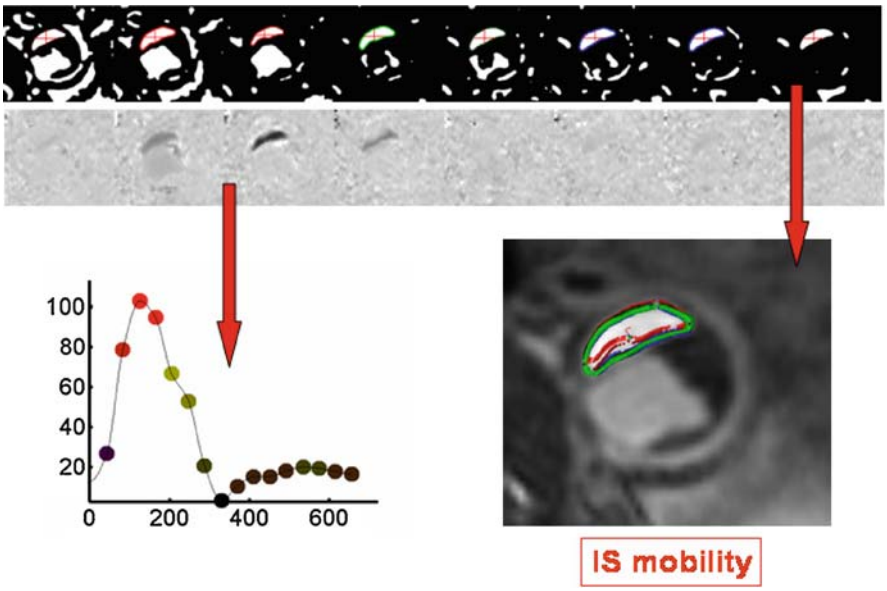


Fig. 5 2D pcMRI images (magnitude images showing anatomical structure on top, flow images showing velocity information below, top) intersecting a TB-AD. Strong flow in the true lumen is noted *after* maximum flow is reached in the false lumen. The boundary of the true lumen is outlined using a color-scheme replicated for displaying the aortic flow waveform at this location (*bottom left*). Overlay of segmented boundaries illustrates IS motion revealing stronger IS mobility at the left compared to the right

In an in-vivo study, Wedding et al. demonstrated the feasibility of cine phase contrast MRI for measuring the velocity of the aortic wall and calculated changes in circumferential strain over the cardiac cycle [27]. Ultrasonometry as the gold standard in in-vitro experiments on a deformable vessel phantom was used to confirm the results [27, 28]. This method was further validated in the porcine thoracic aorta in-vivo model where the motion of implanted markers in the aortic wall was tracked to validate the results obtained with MRI [29].

References

1. Kevles BH (1997) Naked to the bone. Rutgers University Press, New Brunswick
2. Fillller A (2009) The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI and DTI. Nature precedings, vol. doi:10.1038/npre.2009.3267.2: Posted 22 May 2009
3. Hagan PG, Nienaber CA, Isselbacher EM, Bruckman D, Karavite DJ, Russman PL, Evangelista A, Fattori R, Suzuki T, Oh JK, Moore AG, Malouf JF, Pape LA, Gaca C, Sechtem U, Lenferink S, Deutsch HJ, Diedrichs H, Marcos y Robles J, Llovet A, Gilon D, Das SK, Armstrong WF, Deeb GM, Eagle KA (2000) The International Registry of Acute Aortic Dissection (IRAD): new insights into an old disease. JAMA 283:897–903
4. Clouse WD et al. (2004) Acute aortic dissection: Population-based incidence compared with degenerative aortic aneurysm rupture. Mayo Clin Proc 79:176–180
5. Hirst AE Jr, Johns VJ Jr, Kime Jr SW (1958) Dissecting aneurysm of the aorta: a review of 505 cases, Medicine (Baltimore) 37:217–79
6. Eagle KA, DeSanctis RW (1989) Aortic dissection. Curr Probl Cardiol 14:225–78
7. Kamalakannan D, Rosman HS, Eagle KA (2007) Acute aortic dissection. Crit Care Clin 23;779–800, vi
8. Tsai TT, Evangelista A, Nienaber CA, Myrmel T, Meinhardt G, Cooper JV, Smith DE, Suzuki T, Fattori R, Llovet A, Froehlich J, Hutchison S, Distante A, Sundt T, Beckman J, Januzzi JL Jr, Isselbacher EM, Eagle KA (2007) Partial thrombosis of the false lumen in patients with acute type B aortic dissection. N Engl J Med 357:349–359
9. Nienaber CA, Fattori R, Mehta RH, Richartz BM, Evangelista A, Petzsch M, Cooper JV, Januzzi JL, Ince H, Sechtem U, Bossone E, Fang J, Smith DE, Isselbacher EM, Pape LA, Eagle KA (2004) Gender-related differences in acute aortic dissection. Circulation 109:3014–3021
10. Pitt MP, Bonser RS (1997) The natural history of thoracic aortic aneurysm disease: an overview. J Card Surg 12:270–278
11. Akin I, Kische S, Ince H, Nienaber CA (2009) Indication, timing and results of endovascular treatment of type B dissection. Eur J Vasc Endovasc Surg 37(3):289–296. Epub 2009 Jan 25
12. Chemelli-Steingruber IE (2009) Evaluation of volumetric measurements in patients with acute type B aortic dissection–thoracic endovascular aortic repair (TEVAR) vs conservative. J Vasc Surg 49(1):20–28. Epub 2008 Nov 4
13. Feezor RJ, Martin TD, Hess PJ Jr, Beaver TM, Klodell CT, Lee WA (2009) Early outcomes after endovascular management of acute, complicated type B aortic dissection. J Vasc Surg 49(3):561–566; discussion 566–567
14. Lim RP, Shapiro M, Wang EY, Law M, Babb JS, Rueff LE, Jacob JS, Kim S, Carson RH, Mulholland TP, Laub G, Hecht EM (2008) 3D time-resolved MR angiography (MRA) of the carotid arteries with time-resolved imaging with stochastic trajectories: comparison with 3D contrast-enhanced Bolus-Chase MRA and 3D time-of-flight MRA. AJNR Am J Neuroradiol 29:1847–1854
15. Minami H, Sugimoto T, Okada M (1999) Evaluation of acute aortic dissection by cine-MRI. Kobe J Med Sci 45:1–11

16. King KF, Berstein MA, Zhou XJ (2004) Handbook of MRI pulse sequences. Elsevier/ Academic, Amsterdam/New York
17. Weston SJ, Wood NB, Tabor G, Gosman AD, Firmin DN (1998) Combined MRI and CFD analysis of fully developed steady and pulsatile laminar flow through a bend. J Magn Reson Imaging 8:1158–11571
18. Karmonik C, Bismuth JX, Davies MG, Lumsden AB (2008) Computational hemodynamics in the human aorta: a computational fluid dynamics study of three cases with patient-specific geometries and inflow rates. Technol Health Care 16:343–354
19. O'Rourke MJ, McCullough JP (2008) A comparison of the measured and predicted flowfield in a patient-specific model of an abdominal aortic aneurysm. Proc Inst Mech Eng [H] 222: 737–750
20. Scotti CM, Jimenez J, Muluk SC, Finol EA (2008) Wall stress and flow dynamics in abdominal aortic aneurysms: finite element analysis vs. fluid-structure interaction. Comput Methods Biomech Biomed Engin 11:301–322
21. Tokuda Y, Song MH, Ueda Y, Usui A, Akita T, Yoneyama S, Maruyama S (2008) Three-dimensional numerical simulation of blood flow in the aortic arch during cardiopulmonary bypass. Eur J Cardiothorac Surg 33:164–167
22. Steinman DA (2002) Image-based computational fluid dynamics modeling in realistic arterial geometries. Ann Biomed Eng 30:483–497
23. Karmonik C, Yen C, Grossman RG, Klucznik R, Benndorf G (2009) Intra-aneurysmal flow patterns and wall shear stresses calculated with computational flow dynamics in an anterior communicating artery aneurysm depend on knowledge of patient-specific inflow rates. Acta Neurochir (Wien) 151:479–485; discussion 485
24. Karmonik C, Klucznik R, Benndorf G (2008) Comparison of velocity patterns in an AComA aneurysm measured with 2D phase contrast MRI and simulated with CFD. Technol Health Care 16:119–128
25. Karmonik C, Klucznik R, Benndorf G (2008) Blood flow in cerebral aneurysms: comparison of phase contrast magnetic resonance and computational fluid dynamics – preliminary experience. Rofo 180:209–215
26. Karmonik C, Bismuth JX, Davies MG, H. K. Y., Lumsden AB An image processing algorithm for the in-vivo quantification and visualization of septum motion in Type III B – Aortic dissections with cine magnetic resonance imaging, in EMBC 2009, Minneapolis, 2009
27. Wedding KL, Draney MT, Herfkens RJ, Zarins CK, Taylor CA, Pelc NJ (2002) Measurement of vessel wall strain using cine phase contrast MRI. J Magn Reson Imaging 15:418–428
28. Draney MT, Herfkens RJ, Hughes TJ, Pelc NJ, Wedding KL, Zarins CK, Taylor CA (2002) Quantification of vessel wall cyclic strain using cine phase contrast magnetic resonance imaging. Ann Biomed Eng 30:1033–1045
29. Draney MT, Arko FR, Alley MT, Markl M, Herfkens RJ, Pelc NJ, Zarins CK, Taylor CA (2004) Quantification of vessel wall motion and cyclic strain using cine phase contrast MRI: in vivo validation in the porcine aorta. Magn Reson Med 52:286–295

Methodological Advances on Pulse Measurement through Functional Imaging

Thirimachos Bourlai, Pradeep Buddharaju, Ioannis Pavlidis, and Barbara Bass

Introduction

The blood pressure and velocity rise rapidly as a result of the opening of the aortic valve in early systole. This spike in blood pressure and momentum travels the length of the aorta and is passed on to peripheral arteries such as the brachial, the carotid, and beyond. The thus formed pulse is an example of a traveling wave in a fluid medium that involves transport of mass and heat. The alteration of the electric field that moves the heart's muscle and the thermo-mechanical effects of pulse propagation in the vascular network creates opportunities for measurement across different modalities. The method that is considered to be the gold standard for pulse measurement is electrocardiography (ECG) [12]. It produces crisp results because it focuses on the source (heart). Other commonly used methods, such as piezoelectric probing [4], photoplethysmography [13] and Doppler ultrasound [9], focus on the vascular periphery. One main characteristic of all these methods is that they require contact with the subject. There are clinical applications, however, where a contact-free method is desirable. Such applications usually involve sustained physiological monitoring of patients who are in delicate state or form; examples range from sleep studies to neonatal monitoring.

The research presented in this article is in the context of stand-off physiological monitoring through passive imaging, a concept first proposed by Pavlidis et al. [17]. In this context, methods for measuring blood perfusion [16], vessel blood flow [10], breathing rate [15], and pulsation [6, 11, 19] have been developed.

Specifically, regarding pulsation, Chekmenev et al. [6] developed an interesting thermal imaging method that used wavelet analysis to quantify pulsation. Good performance results were reported on a dataset of eight subjects. The issues of tissue tracking and sensitivity analysis, however, were not adequately addressed.

T. Bourlai (✉), P. Buddharaju, and I. Pavlidis
Computational Physiology Lab, University of Houston, 4800 Calhoun Rd, Houston, TX, USA
e-mail: tbourlai@uh.edu

B. Bass
Department of Surgery, The Methodist Hospital, 7111 Fannin St, Houston, TX, USA

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_6,

Garbey et al. [11] and Sun et al. [19] developed different thermal imaging pulsation methods that used Fourier analysis. The dominant heart rate frequency was estimated by averaging the power spectra of each pixel in a preselected segment of a superficial vessel. Two variant methods were developed: the Along (ALM) [11] and the Across (ACM) [19]. The thermal imprint along (ALM) or across (ACM) the center line of a large superficial vessel was selected. Both methods were limited by the use of a non-optimal parameter set, the presence of tracking errors, and the lack of in-depth statistical analysis.

This paper addresses the limitations of previous contact-free approaches of pulse recovery and reports substantial methodological advances. The new pulse recovery thermal imaging (PRETI) method features parameter optimization for both the ALM and ACM models, which it uses within a fusion scheme. In this fusion scheme, there is a choice between three tissue tracking algorithms. These are the coalitional tracker, where a single tracking network is used, the tandem tracker, where two tracking networks are used in conjunction, and the micro-tracker, which features fine-tuning capability. The investigation reveals which measurement model (ALM or ACM) can pair with what motion tracking algorithm to offer a better trade-off between performance and computational complexity. All combinations are compared with the baseline (REF) pulse measurement methods [11, 19] in a set of experimental data drawn from 12 subjects, per approved protocol from the local Institutional Review Board (IRB).

The rest of the paper is organized as follows: section "Pulse Measurement Methodology" describes the new measurement methodology. Section "Experimental Setup" describes the experiments. Section "System Optimization" presents the optimization results. Finally, section "Conclusions" concludes the paper.

Pulse Measurement Methodology

PRETI is a fusion scheme that involves five steps: (1) Selection of Region of Interest (ROI); (2) Motion Tracking (tracking the ROI by using the single, sequential, or automatic tracker); (3) Blood Vessel Registration; (4) Noise Cleaning; and finally (5) Statistical Analysis. Figure 1 illustrates the steps of the new methodology that conclude with the computation of pulse. The computed pulse is compared against a "ground-truth" measurement provided by an ADInstruments piezoelectric device [1].

Selection of Region of Interest (ROI)

For a successful measurement of cardiac pulse via thermal imaging, selection, and knowledge of the anatomical region of interest is important. Stand-off pulse measurements are typically performed on the face, because it is easily accessible and

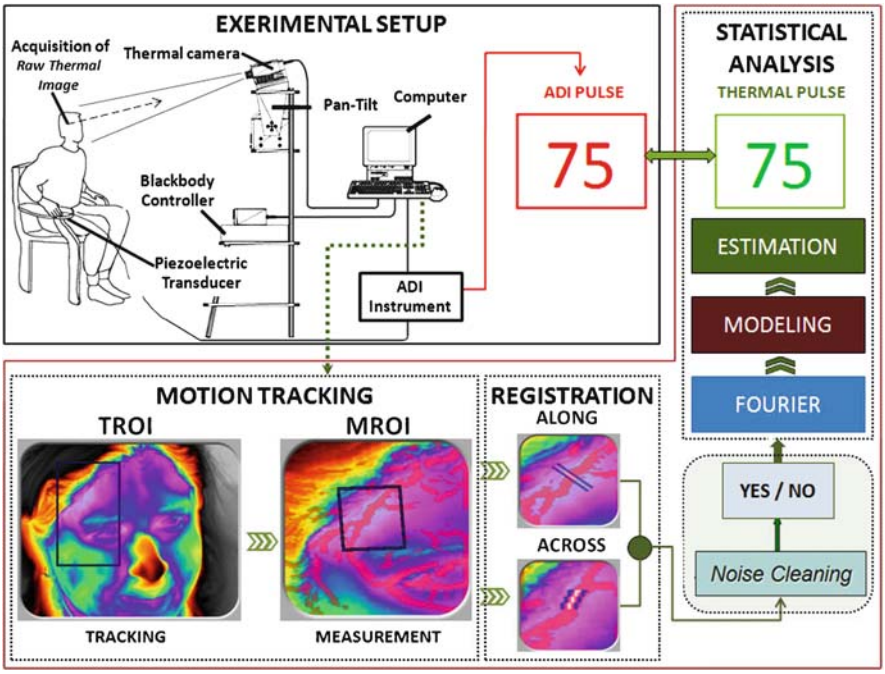


Fig. 1 Outline of the new pulse measurement methodology. The thermal imaging measurements are compared against "ground-truth" values provided by an ADInstruments piezoelectric device [1]

features major superficial vasculaturization. Periodic pulsation in facial vasculature results in localized skin temperature modulation through the mechanism of thermal diffusion [10].

Most of the facial vasculature is derived from the external carotid artery (ECA). The superficial temporal artery (STA) is a terminal branch of ECA. STA begins between the ear and the Temporo–Mandibular joint (TMJ) ascending upward and eventually splitting in the upper head area into the frontal and parietal branches (Fig. 2).

Pinar and Govsa [18] reported an excellent study on STA anatomy, its arterial branches, and their importance. For the purpose of thermal pulse measurement, STA is the region of choice because it is the most superficial vessel on the face and still has substantial size (2.73 ± 0.51 mm).

Motion Tracking

The proposed pulse measurement method is contact-free; hence, in the absence of good tracking, even the slightest movement by the subject will shift the ROI from its initial selection. PRETI employs three tracking algorithms, which are described

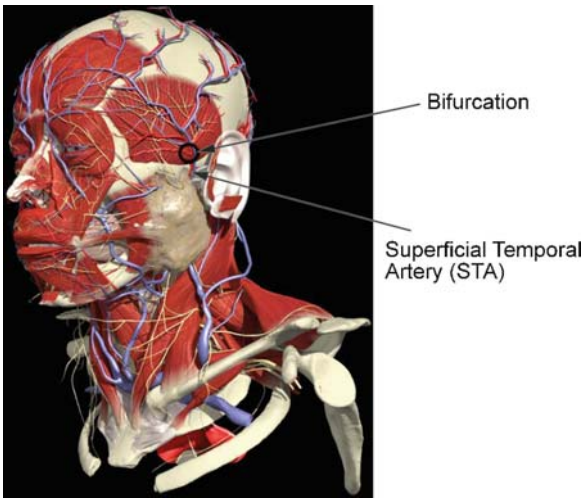


Fig. 2 Illustration of the superficial temporal artery (STA) and its bifurcation around the zygomatic arch – from Primal Pictures [14]

below, to compensate for motion and provide consistent ROI measurements over time. At this point, it is important to clarify that there are two types of ROI: The broader ROIs, which are the tissue areas upon which the trackers operate – these are the tracking ROIs (TROI). The smaller ROIs that are within the TROIs and atop the vessel's thermal imprint, where the measurement is performed – these are the measurement ROIs (MROI).

Coalitional Tracker

The coalitional tracking algorithm [7] optimizes collaboration among many simple particle-filter trackers, to achieve robustness and precision usually attainable only by model-based trackers. It was explicitly developed to support reasonable accuracy of vital sign measurements in thermal infrared, without resorting to modeling. Please note that modeling of highly dynamic imagery of physiological function is quite difficult. The coalitional tracker's performance deteriorates in the presence of out-of-plane rotations, which are due to pose changes of the subject's face. Figure 3 clearly illustrates such a case. The change in facial pose from time $t = 0$ (Fig. 3a) to time $t = 10$ sec (Fig. 3b) has caused the coalitional tracker to loosen its grip on TROI. At every point in time, the new TROI, as determined by the coalitional tracker, is used to produce the new MROI through a geometric transformation fixed during initialization. Thus, a small TROI error unavoidably translates to a small MROI error – typically a few pixels. Unfortunately, the thermal footprint of the vessel is also just a few pixels wide. Thus, even small tracking failures can throw MROI outside the vessel's thermal footprint and introduce substantial measurement errors (Fig. 3c, d).

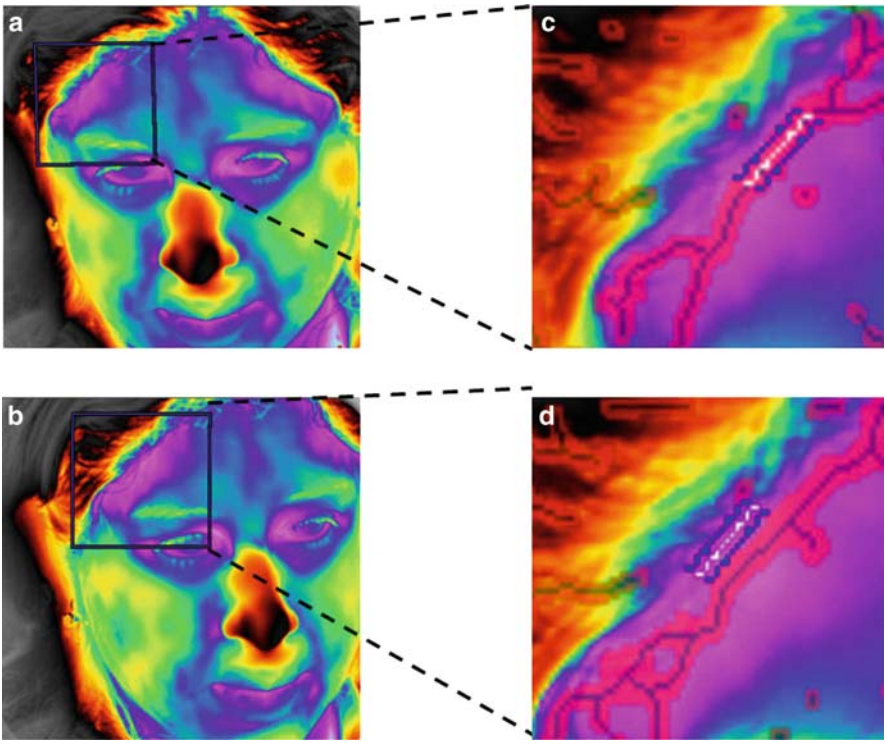


Fig. 3 Example of coalitional tracker performance. Thermal snapshots of the subject at time (**a**) $t = 0$ and (**b**) $t = 10$ sec with a coalitional tracker targeting the temporal area. Blow-ups of MROI at time (**c**) $t = 0$ and (**d**) $t = 10$ sec, where drift is evident. The tracker cannot cope effectively with pose changes

Tandem Tracker

The tandem tracking algorithm uses two coalitional trackers to overcome the errors introduced by a single coalitional tracker. The top coalitional tracker tracks a large TROI, which is centered in the general temporal area. At each point in time, this tracker provides cue about the initial position of another coalitional tracker, which tracks a smaller TROI centered on the temporal vessel. This "inside" coalitional tracker performs its own local tracking, using the cue from the top tracker as an initialization. Finally, the second tracker determines the position of the MROI, through a geometric transformation set at the beginning. Large coalitional trackers are prone to drift (and small inaccuracies), while small coalitional trackers to loss (and total failure). However, if small coalitional trackers do not get lost due to abrupt motion, they can afford much more accurate tracking than large ones. The tandem coalitional tracking scheme capitalizes upon these complementarities to deliver optimal performance (see Fig. 4).

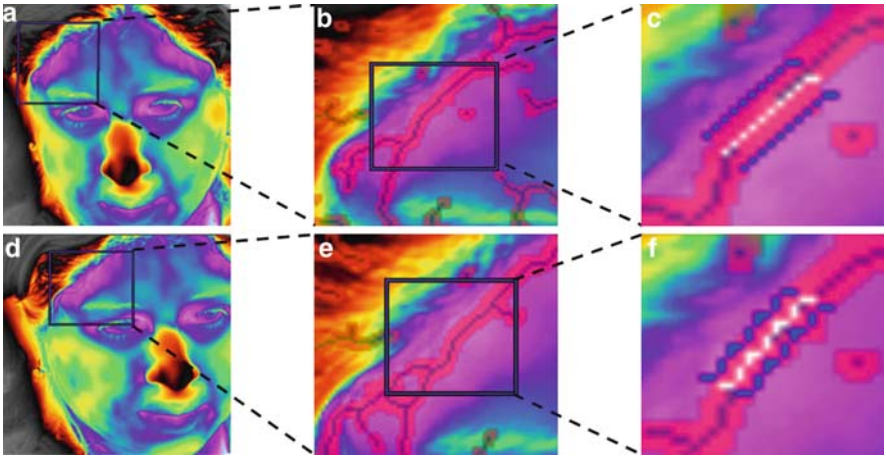


Fig. 4 Example of tandem tracker performance. Thermal snapshots of the subject with large TROI on the general temporal area, small TROI around the temporal vessel, and MROI on the thermal imprint of the temporal vessel at time $t = 0$ (**a**–**c**) and $t = 10$ sec (**d**–**f**)

Microtracker

The tandem tracker performs better than the single coalitional tracker at a premium computational cost. As a way to improve performance but at a more moderate computational cost, a third (and novel) micro-tracking algorithm was introduced. This algorithm, uses a single coalitional tracker in the general temporal area, which does not control the relative position of MROI through a rigid geometric transformation. Instead, MROI is localized at each point in time through a segmentation algorithm that operates in a smaller area around the temporal vessel. This is reminiscent of the inner TROI in the tandem tracker. The difference is that instead of secondary tracking, segmentation is taking place here. This segmentation fine-tunes the localization of the vessel's thermal imprint, stopping in essence the propagation of error from the top tracker (Fig. 5). The segmentation process involves the following steps:

Step 1: Within the inner ROI, perform top-hat segmentation to differentiate the vessels from the surrounding tissue [5]. The vessels' thermal imprints are usually at a gradient from the remaining region, because of convection from the flow of hot arterial blood.
Step 2: Thin the blood vessel network down to one pixel thickness [5].
Step 3: Find the largest vessel in case there are more than one within the inner region.
Step 4: Find the best fit for the points of the largest vessel through linear regression.

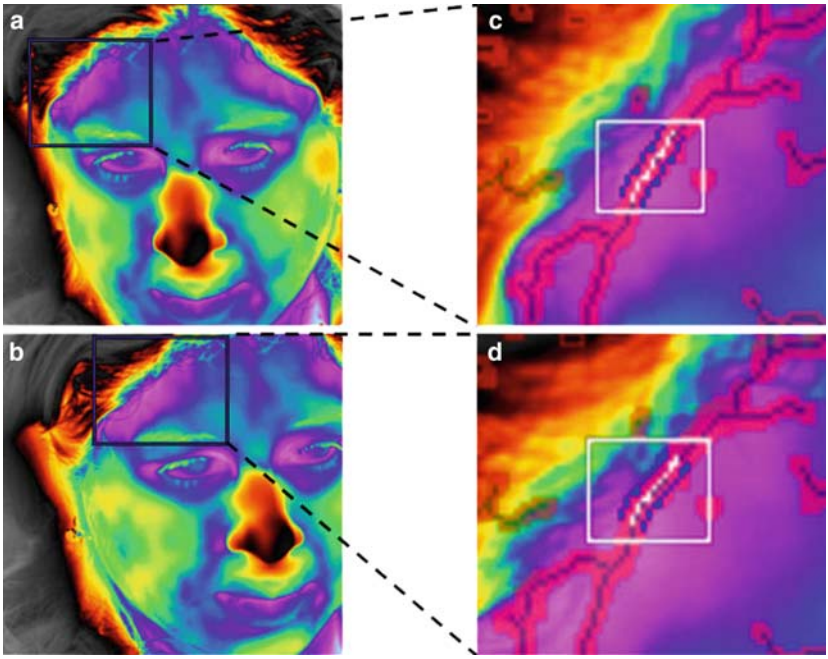


Fig. 5 Example of a microtracker performance. Thermal snapshots of the subject at time (**a**) $t = 0$ and (**b**) $t = 10$ sec with a coalitional tracker targeting the temporal area. Blow-ups of MROI and the inner region, where vessel segmentation is taking place, at time $t = 0$ and $t = 10$ sec

Blood Vessel Registration

The operator can select a thermal imprint along the center line of the STV or at 90° across it applying the ALM or ACM model respectively. The algorithm expands symmetrically into an elongated rectangle. The width of this rectangle depends on the width of the STV on the thermal imagery. For a subject imaged at 6 ft with a 50-mm lens, and when using the ALM model the rectangle's width is 3–7 pixels. By convention, in the ALM model, we place the x-axis of our coordinate system along the width and the y-axis along the length of the FSTA. We have the opposite coordinate system in the case of the ACM model. Note that the use of a rectangle in the case of ACM is a new approach investigated in our experiments. In the previous version of our ACM model, only a single pixel line of 90° across the vessel under study was considered.

1. *Tracking noise cleaning algorithm (TNCA)*. In the next stage, our newly developed TNCA process can be selected. It is a three-stage algorithm that assists our tracker in the selection of high confidence frames and by correcting the overestimated maximum pulse frequencies. Here follows a description of each stage:

(a) *Tracking confidence estimate (TCE)*. In order to achieve robust tracking, we have developed an algorithm that utilizes the final stage (target estimation) of the coalitional tracking process to decide whether a frame should be included in the pulse measurements or not. During target estimation, the winning coalition is produced by using a weighted sum of 4 scores at time t: Template Match, Geometric Alignment, Inter-frame projection agreement; and Inter-frame membership retention [7]. Confidence is the average template matching score (a number between [0, 1]) of all the trackers in the winning coalition. This score is computed frame by frame and rewards trackers that maintain consistent imaging content. The decision to include the frame at time t to the pulse measurement estimation process is based on a 70% confidence threshold. An example case is presented in Fig. 6.

(b) *Temperature thermal imprint estimate (TTIE)*. TTIE further assists the tracker to select frames where the segmented thermal imprints of the STV are of high quality. The thermal imprint at frame $t-1$ (TIF_{t-1}) is compared with the current thermal imprint at frame t (TIF_t). The comparison is based on the complement of the absolute normalized difference (CAND)$(1-\mathrm{ABS}(\mathrm{TIF}_t-\mathrm{TIF}_{t-1})/\mathrm{TIF}_{t-1})$, which is the absolute difference between the TIF_t and TIF_1 measurement normalized against the TIF_{t-1} and subtracted from unity. This gives a weighted indication of how close the TIF_1 measurement is to the TIF_{t-1} measurement in each case. Again, the decision

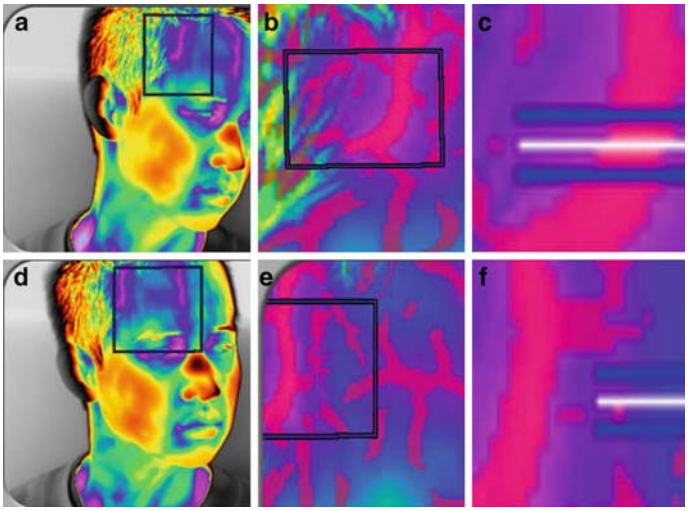


Fig. 6 Example cases of using the tracking confidence (TCE) and temperature thermal imprint estimate (TTIE) noise cleaning steps. (**a**) We can see the TROI where the tracker is having an 80% confidence, the resulting MROI at (**b**), and STV registration at (**c**) where the TTIE is 95%. (**d**) The tracking ROI where the tracker is having a 55% confidence, the resulting measurement ROI at (**e**), and frontal STV registration at (**f**) where the TTIE is 18%. In the (**d**), (**e**), and (**f**) cases the associated frame is rejected

to include the current frame to the pulse measurement estimation process is based on a 95% confidence threshold. Example cases where the TTIE is 95% vs. 18% are presented in Fig. 6c, f, respectively.

(c) *Peak correction*. Our experiments are set up in a quiet indoor environment and we test healthy subjects who are relaxed during the recording. Under these conditions, it is reasonable to assume that their pulse should range between 40 and 100 beats per minute (bmp). Therefore, we can facilitate pulse recovery by removing signals with frequency lower than 0.67 Hz (40 bmp) and higher than 1.67 Hz (100 bmp). This pulse range is selected by setting the low/high Pulse values in the common parameters section of the UI. However, there is still a possibility to overestimate/underestimate the maximum frequencies computed during pulse estimation. These can be triggered by a low/high estimation of the camera frame rate that affects the conversion of the pulse frequency from BPM to Hz (see (1)).

$$\text{Low/High rate} = \frac{\text{Low/High Pulse(BPM)} \times \text{TimeWindow}}{\text{Sec\_Per\_Min} \times \text{FrameRate}} \text{(Hz)} \quad (1)$$

Even though it is computed to be on average 30 frames per second (fps), it may go even below 20 fps or even above 45 fps. Thus, we may recover a sudden and very high/low pulse (e.g. 120/30 bpm) of a normal subject with an average pulse of 70 bpm. We know that in long observation periods the pulse frequency is expected to dominate in the spectral domain since it is more consistent than white noise. Therefore, in such cases, the algorithm can replace the overestimated pulse or peak frequency with the dynamic mean pulse frequency (DMPF) (see step three of the ACM model in Appendix).

DMPF is initially computed over an extended period of time T ($T \geq 30$ s). In our new methodology and in order to achieve a better DMPF measurement, we keep updating the mean pulse measurement. The estimation starts from the beginning of the clip and can be updated after each time window of 2^N frames ($N \in [7, 11]$) or $4k$ s ($k \in [1, 5]$). A typical window selected by the user through the UI is 64 frames or ≈ 0.5 s.

An example case where the peak correction step is employed is presented in Fig. 7.

2. *Statistical analysis*. In this step, we apply a Fourier-based method on the tracked ROI of the STV. It is applied in a novel manner to capitalize upon the pulse propagation effect and extract the dominant pulse frequency. Two models are used, the ALM and ACM when we select to operate either along or across the STV. By operating on the frequency domain and combining appropriately the power spectra of the time evolution signal of the temperature profiles, the signal can be reinforced. Thus, in the next stage, the adaptive estimation filtering (AEF) is employed in the same manner after either the ALM or the ACM modeling. AEF convolves the FFT outcome with a normalized historic power spectrum. In the last step, the cardiac pulse is computed by recovering the highest energy content of the filtered power spectrum.

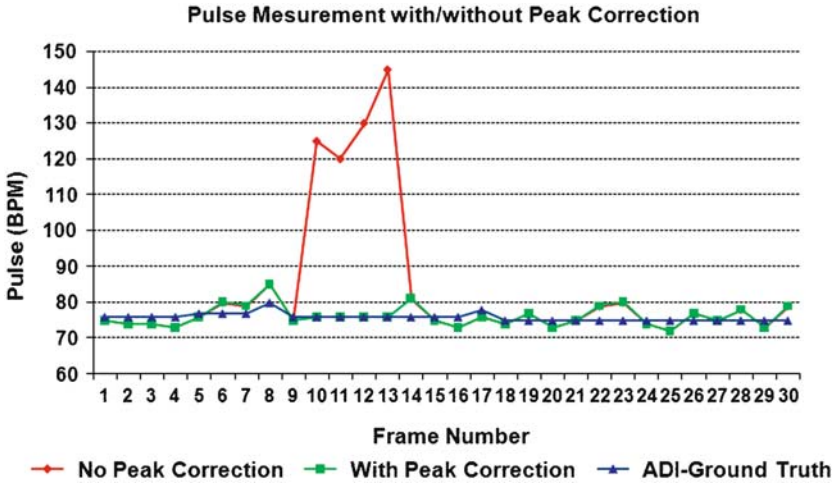


Fig. 7 Peak Correction example case. The algorithm can pick up wrong estimations of the thermal pulse at time t and correct them with the latest mean pulse update

A brief description of the ALM, ACM and AEF can be found in the Appendix. More details can be found on [11, 19].

Experimental Setup

A high-quality TI system has been designed for data collection that can obtain highly accurate measurements. The centerpiece of the system is an MWIR camera [2] with 50-mm lens. The lenses allow focusing on parts of the subject with rich superficial vasculature (e.g. face) at distances between 3 and 10 ft. The camera is capable of capturing 30 fps in a full spatial resolution of 640 × 480 pixels. The camera is placed atop a pan-tilt device to allow flexible positioning. We also use a differential blackbody as a calibrating device. The temperature resolution of the blackbody matches that of the thermal camera. To achieve maximum portability, all the aforementioned hardware components are placed in a cart and communicate with a powerful workstation.

Data collection is performed using healthy subjects. Subjects suffering from neuropathies, micro or macroangiopathy, as well as strong smokers have been excluded. Before data collection the subjects are briefed, and after that they signed a consent document. During data collection, each subject is placed about 6 ft away from the TI system. Then, as reported in section "Pulse Measurement Methodology," the TI measurements computed by our system are compared with the GT measurements. These are reported by the piezoelectric device. The MLT 1010 piezoelectric pulse transducer used is wired to the subject's index finger tip. Our pulse measurement experimental setup can be viewed in Fig. 1.

System Optimization

The optimization framework of the PRETI system in terms of performance and computational cost is quite complex. Global optimization requires an exhaustive evaluation of an uncountable number. Hence, in practice, only partial optimization is feasible with many parameters taking default values after an efficient parameter selection process. The main design issues are to fine-tune the harmonic analysis of the signals through parameter optimization of the baseline models, and then to improve the quality of the extracted physiological signals through sophisticated tracking and a noise reduction algorithm.

To evaluate the performance of our PRETI system, we employ three optimality criteria. The first and most important criterion is the Paired Student's t test (PSTT), a statistical hypothesis test that is used to compare two sets of quantitative data (in our case ground truth pulse and pulse estimation data). Please note that in [20] we have also used the Wilcoxon signed-rank test, as an alternative to the paired Student's t-test since the pulse measurements cannot be assumed to be normally distributed. However, the trend of the results did not change. Here we will report only the t-test results. We also calculate the cumulative sums (CUSUM) between the instantaneous pulse measurements and their corresponding ground-truth ones. In (2), we can see how the cumulative percentage error for subject i is computed.

$$E^i_{\text{cum}} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|S^i_{\text{T}}(t) - S^i_{\text{G}}(t)\right|}{S^i_{\text{G}}(t)} \tag{2}$$

Finally, we compute the normalized mean pulse difference (NMPD) which is based on the CAND of the mean pulse against the mean ground-truth measurements over the whole thermal clip (see (3)).

$$\text{NMPD} = 1 - \frac{\left|\overline{S}_{\text{T}} - \overline{S}_{\text{G}}\right|}{\overline{S}_{\text{G}}} \tag{3}$$

We tested our approach on 12 subjects (6 male/female) with ages in the range of 24–55 years old. The description of the parameter selection process follows.

Parameter Selection

The two main goals of this investigation are first to find the most efficient parameter set in terms of performance for each model and for all subjects. Then we prove that the application of those sets in combination with the use of the TNCA algorithm provide an additional performance advantage. Based on previous studies and the physiology of the vessels of the subjects under study, the baseline system (REF) involves both the ALM and ACM models and the fixed parameters are the length

(L) of the thermal imprint (7–10 pixels), the pulse range (40–100 bpm), and the time window (512 frames). In the REF system, there is no TNCA algorithm employed.

In this study for both models, we keep the length (L) of the thermal imprint and the pulse range that we restrict our investigation fixed as before. We investigate further into the importance of the width (W) of the thermal imprint, and the time window (frame range) of the history data. What follows is the investigation performed to identify the optimum parameter set per model in terms of performance.

- *ALM*. In the case of the ALM model, the values of W investigated are extended from 1 to 11 pixels in six steps $(1, 3, 5, \ldots, 11)$. The choice of these prime numbers guarantees that the central line selected by the operator is always in the middle on the vessel and the additional pixel lines are to the right and left of it. Also this range is practically selected so that any vessel width in the forehead of our 12 subjects is included within the MROI Fig. 8.
- *ACM*. In the case of the ACM model, the range of W is extended from 1 to 13 pixels in seven steps $(1, 3, 5, \ldots, 13)$ as shown in Fig. 9. Some additional parameters that are investigated are the quadratic interpolation (to apply or not and how many times) as well as the mean and variance of the normal distribution $N(\mu_p, \sigma_p^2)$.
- *Both models*. In both of our models, we extend the investigation of the time window from 128 and up to 2,048 frames (2^N frames for $N \in [7, 11]$) and we include the use of the TNCA algorithm and its associated parameters.

In Table 1, we present the parameters used before and after optimization.

The results of applying the optimum parameters in the REF system are presented in Table 2 where we can compare the performance of the baseline system with the optimized one, with and without employing the TNCA algorithm. Note also that

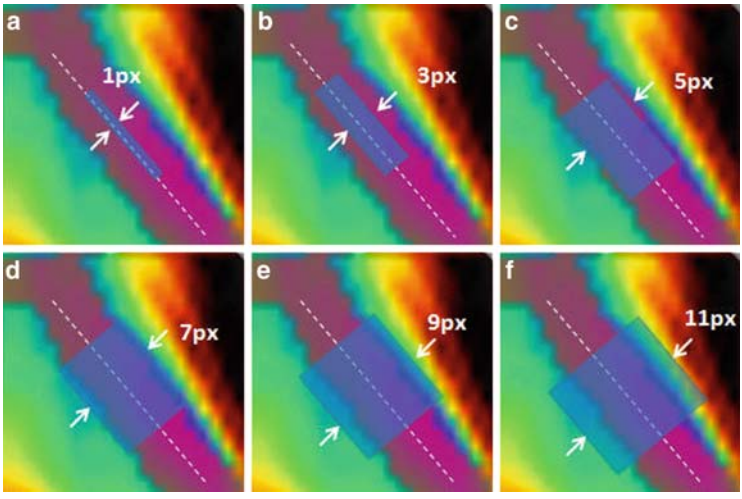


Fig. 8 ALM : Example case where the vessel width is 7 pixels. We can see how the ALM algorithm performs when we select different rectangles along the vessel orientation. The widths selected are (**a**) 1 px, (**b**) 3 px, (**c**) 5 px, (**d**) 7 px, (**e**) 9 px, and finally (**f**) 11 px

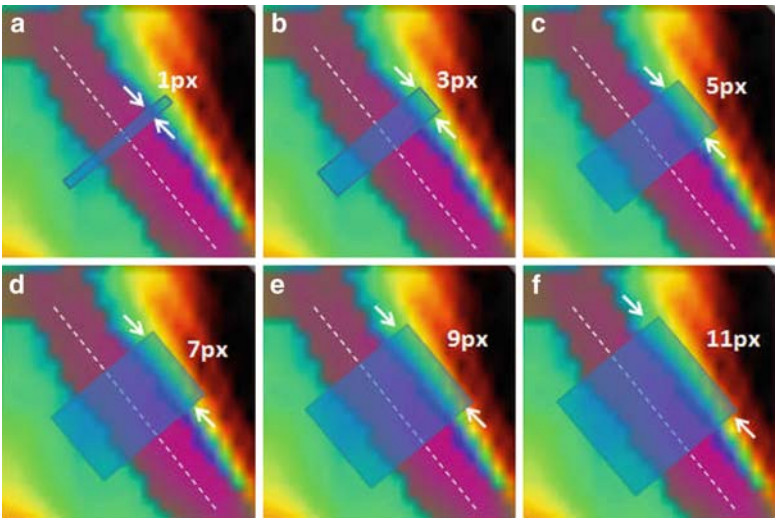


Fig. 9 ACM : Example case where the vessel width is 7 pixels. We can see how the ACM algorithm performs when we select different rectangles vertical to the vessel orientation. The widths selected are (**a**) 1 px, (**b**) 3 px, (**c**) 5 px, (**d**) 7 px, (**e**) 9 px, and finally (**f**) 11 px

Table 1 Parameters used before and after optimization

Models	Opt	TNCA	Length (px)	Width	TW	QI(ta)
ALM	No	No	7–10	3–7 px	512	–
	Yes	Yes	7–10	1 px	128	–
ACM	No	No	7–10	1 px	512	5
	Yes	Yes	7–10	7	2,048	1

Opt optimization, *TW* Time Window, and *QI(ta)* Quadratic interpolation (Times applied)

Table 2 Performance results before (REF) and after an efficient parameter selection and the use of the TNCA algorithm (OPT)

		REF			OPT		
MDL	TNCA	TT	CuSum	NMPD	TT	CuSum	NMPD
ALM	No	Fail	10.54	88.63	Fail	7.78	97.26
	Yes	–	–	–	Pass	6.89	97.57
ACM	No	Fail	10.14	94.56	Fail	8.15	98.50
	Yes	–	–	–	Pass	6.88	98.55

OPT Optimization, *REF* Baseline, *MDL* Model, *T T t* test

while the optimized parameter set was achieved for each subject, due to space constraints, here we present only the average performance results for all subjects.

Based on the above results, we can see that the efficient parameter selection achieved better performance results and that the use of the TNCA algorithm is the key for the statistical test to pass. Hence, in the next step, we select the optimum pa-

rameter set per model, we apply the TNCA algorithm, and then we investigate which motion tracking methodology we should follow to achieve maximum performance with the minimum computational cost. There are three alternative tracking optimization approaches that we can use, i.e. the SITA, the SETA and the AUTA as described at Section "Introduction." These are compared to the REF system. This study is analyzed in section "Motion Tracking."

Motion Tracking

After the optimum parameter selection for each thermal signal analysis model, our goal is to investigate various testing configurations. In our design strategy, we investigate the three alternative PRETIS tracking optimization approaches, the SITA, the SETA and the AUTA. By employing the optimum parameters and TNCA algorithm, we performed ten experiments per tracking approach for each subject. Then, the results are averaged and finally compared with the baseline approach. We use the same optimality criteria as described in the beginning of section "System Optimization," and at the same time, we compute the response time for each configuration. Both the performance as well as the response time results are presented below.

Performance Results

In Figs. 10–12, we illustrate the performance results for each subject when employing either the ALM or ACM model with the use of the optimum parameter set as well as the TNCA algorithm identified earlier. We can see the results after running each experiment 10 times for each of the SITA, SETA, or AUTA motion tracking approach. The criteria used are the CuSum and the NMPD as described in the beginning of section "System Optimization."

In Table 3, the final performance results are presented when averaging the above results for all subjects. We can see the benefits of optimization when comparing the REF and SITA approaches where only a single tracker was used before and after parameter optimization while using also the TNCA algorithm. Based on the t test criterion, it is clear that the SETA approach offers by far the best results. The CUSUM and NMPD criteria also support our selection. Finally, note that the AUTA approach did not perform as well as expected. Although very good results have been achieved, micro-tracking needs to be further optimized. This is planned to be part of our future work.

Response Time Results

The response time results are presented in Fig. 13 when using the REF system or any of the other three motion tracking optimization approaches. We are using blue color for the ALM model and red color for the ACM model. With blue is also indicated

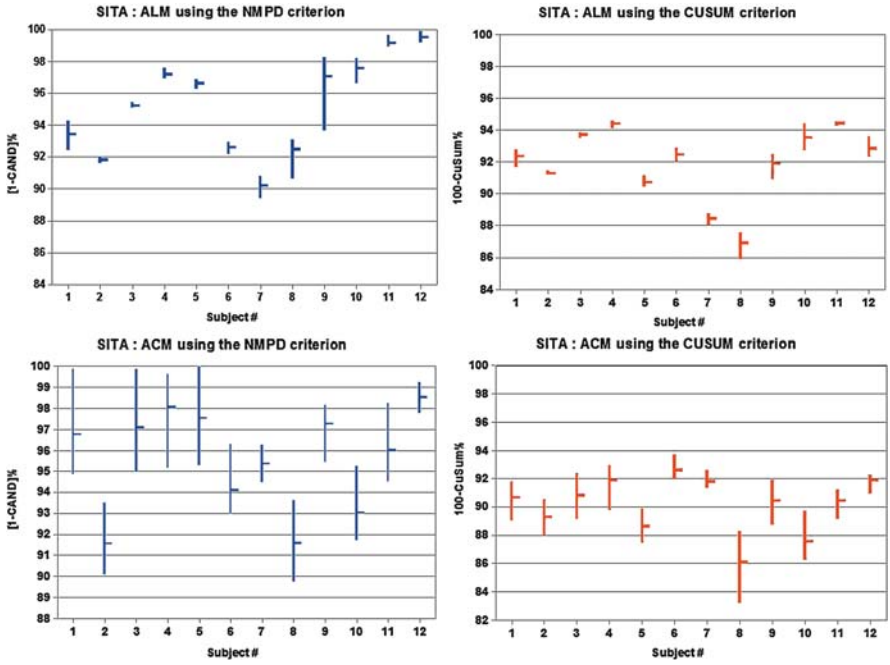


Fig. 10 ACM with SITA: Performance results when using the CuSum and the NMPD criteria. We perform ten experiments per parameter set and highlight the min, mean, and max values per subject and for each criterion employed

the average video time of all subjects that are processed in this study. In that way, we can see whether any of our approaches can process real-time all frames captured by the thermal camera.

In terms of time, the SITA approach when using the optimized ALM model gives the best time. However, in combination with our performance results, we highlight as our best option the SETA approach when using the optimized ALM model.

Conclusions

In this paper, we report substantial improvements in the design and methodology of a baseline pulse recovery thermal imaging system. The experimental results demonstrate that a proper parameter selection and the use of a TNCA algorithm have improved considerably the performance of our proposed PRETIS system when compared with the baseline one. TNCA in particular improves system performance up to 15.6% while imposing only a maximum of 3.98% increase in the computational complexity of the system.

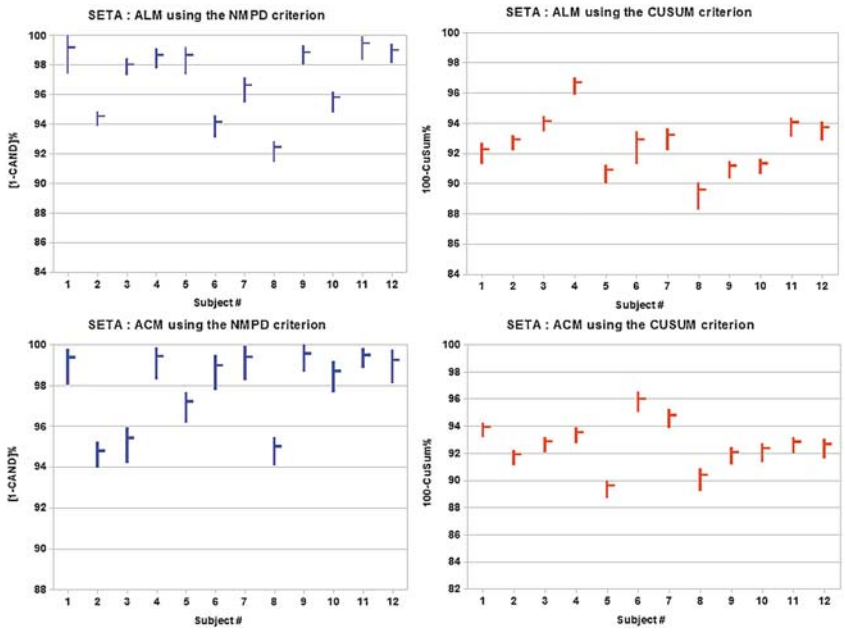


Fig. 11 ALM/ACM with SEQTP: Performance results when using the CuSum and the NMPD criteria. We perform ten experiments per parameter set and highlight the min, mean, and max values per subject and for each criterion employed

After an efficient parameter selection in either model of the baseline system, we compare three motion tracking methodologies in terms of system performance and response time. When using either model, the response time of SITA is the lowest one when compared with all other approaches and when using as a reference the baseline approach. However, it failed in almost 84% of the statistically significant tests. Even though SETA is 13% slower than SITA, it is still 11% faster than the baseline system when the ALM model is employed. SETA when using the ALM model proves to be the best overall solution also in terms of performance since it passes all statistically significant tests, achieving also the lowest CuSum error (7.26%) and the highest accuracy in terms of NMPD (97.11%). Similar results are achieved when using the SETA approach with the ACM model (almost all t tests passed, CuSum 7.27%, NMPD (98.04%). However, in this case, the response time is at least 4 times higher than when using the ALM model and thus it is considered to be our second best choice. Finally, even though AUTA was designed to further improve our pulse estimations by minimizing noise, it fails in about 80% of the statistically significant tests and thus its operation requires some further investigation.

Investigating different optimization strategies on a PRETI system is an interesting task. However, the conclusions of this work were drawn in the context of the our database. For future work, we plan to perform a new data collection process where a minimum of 30 subjects will participate. A new camera with a

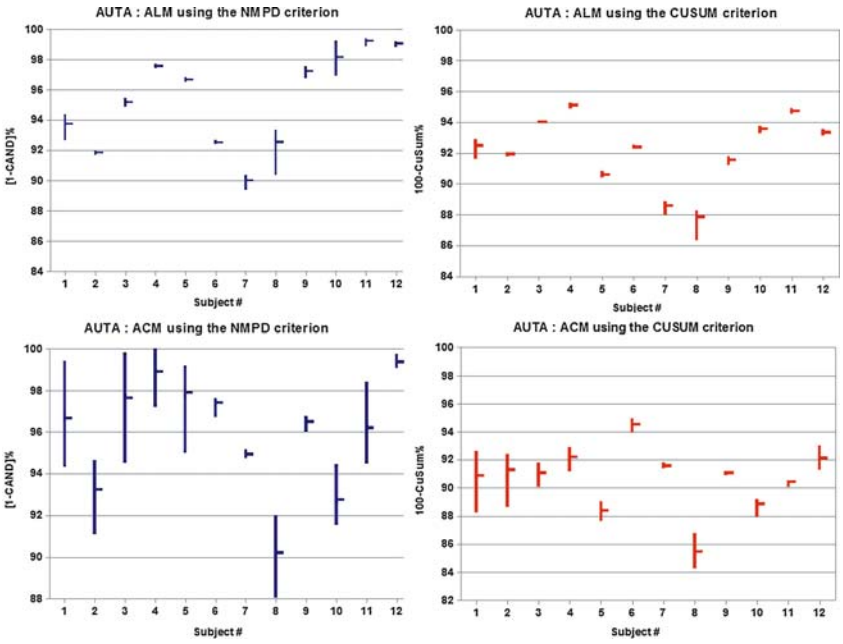


Fig. 12 ALM/ACM with AUTA: Performance results when using the CuSum and the NMPD criteria. We perform ten experiments per parameter set and highlight the min, mean, and max values per subject and for each criterion employed

Table 3 Final performance results when using the REF system and all our motion tracking approaches after optimization

Criteria	Method		REF	SITA	SETA	AUTA
T test	ALM		0	12	100	10
	ACM		0	17	97	20
CUSUM	ALM	Mean	10.56	8.11	7.26	7.82
	ACM	STD	10.14	9.84	7.27	9.35
	ALM	Mean	1.249	0.948	0.585	0.003
	ACM	STD	1.374	1.106	0.581	0.008
NMPD	ALM	Mean	88.63	95.24	97.11	95.32
	ACM	STD	94.56	95.23	98.04	95.04
	ALM	Mean	0.743	0.563	0.682	0.422
	ACM	STD	2.243	1.752	0.683	1.691

The mean values and standard deviation for all subjects are presented after running each experiment 10 times per subject
Opt Optimization, *REF* Baseline

better spatial resolution and a new protocol will be used that has been designed so that motion noise is minimized. Furthermore, we intend to optimize separately our new AUTA approach and to design a new theoretical framework that will improve system performance. We believe that our advanced research work will find

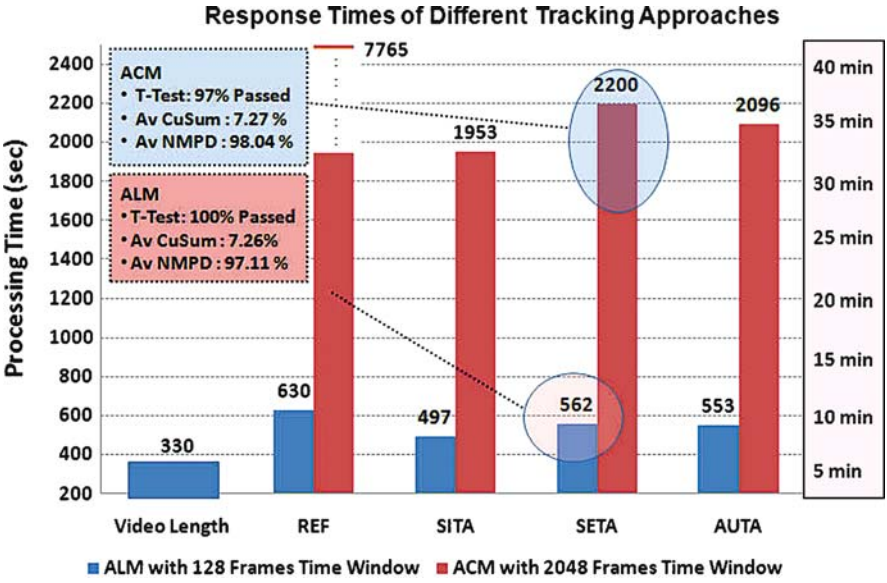


Fig. 13 Response time results when using the three motion tracking approaches. The best cases in terms of performance are also highlighted

great applications in the areas where the monitoring of heart rate through a passive sensing system is needed and in the cases where motion artifacts and poor subject cooperation are considered a serious problem.

Acknowledgements Research activity involving human subjects has been reviewed and approved by the University of Houston Committee for the Protection of Human Subjects. The authors would like to thank all the volunteer subjects who participated in their test population. They would also like to thank Dr. E. Glinert from the National Science Foundation (NSF) for his support and encouragement in this nascent technology effort. Equally, they would like to thank Dr. J. Levine from the Mayo Graduate School of Medicine for his valuable feedback.

Appendix

Description of the ALM Model: Here follows a brief description of the Along the Vessel Model:

- *In step one*, within the MROI, the operator selects manually or automatically a straight segment of 7–10 pixels (depending on the vessel selected) along the center line of the superficial blood vessel. The algorithm expands symmetrically into an elongated rectangle the width of which can be from 1 to 13 pixels (as opposed to 3–7 pixels used before). The width of this rectangle depends on the width of the vessel on the thermal imagery.

- *In step two*, we record the time evolution of the pixel matrix delineated by rectangle R for 2^N frames, where $N \in [7, 11]$ (only the use of $N = 9$ or 512 frames was reported in our previous studies. In this paper, we investigated all values of N). Thus, we produce a 3-D matrix $A(x, y, t)$, where x and y is the spatial extent of rectangle R and t is the timeline.
- *In step three* and in order to reduce the noise, we average the pixel temperatures along the x dimension.
- *In step four*, for each effective pixel on the measurement line, we obtain the time evolution signal of its temperature. We apply the FFT on each of these signals.
- *In step five*, we average all the power spectra computed in the previous step into a composite power spectrum.

Description of the ACM Model: Here follows a brief description of the Across the Vessel Model:

- *In step one*, the operator draws manually or automatically a line that traverses the cross section of the thermal imprint of the vessel (e.g., FSTA). The section spans between 1 and 15 pixels (as opposed to 3–7 pixels used before). The spatial resolution of the measurement line is increased by applying quadratic interpolation once (as opposed to 5 times used before) to minimize the computational complexity while achieving good performance. We model the cross section temperature function using the first five (5) cosine functions of the Fourier series.
- *In step two*, we compute the ridge and the boundary points at each frame. The first corresponds to the middle of the vessel's cross section, where the blood flow speed is maximal, while the second is recorded at the vessel's boundary where the minimum blood flow speed occurs. The time evolution of these points form the ridge and boundary temperature functions (R_{TF} and B_{TF}) respectively.
- *In step three*, we compute the static mean pulse frequency (SMPF). We apply the FFT on both R_{TF} and B_{TF} 1D signals and obtain their power spectrum (P_{r} and P_{b}). We model both power spectrum as a multi-normal distribution by applying a Parzen window method [8] and get the multi-normal distributions P'_{r} and P'_{b}. We multiply P'_{r} and P'_{b} to obtain the combined model spectrum P'_{rb}. Then, we find the frequency f_n for which P'_{rb} assumes its maximum amplitude. The f_n frequency is considered the SMPF of the subject during the time period of the first $T \geq 30$ s or $\approx$ 1,024 frames and it is represented as the normal distribution $N(\mu_{\mathrm{p}},\sigma^2_{\mathrm{p}})$ with mean $\mu_{\mathrm{p}}=f_{\mathrm{p}}$ and variance σ^2_{p}.

  In this paper, we go a step further and compute the dynamic MPF (DMPF). This is performed by updating the MPF for every 64 frames after the first 1,024 frames. We have also optimized the value of the variance to achieve better performance results.
- *In step four*, we compute the instantaneous pulse frequency (IPF). We apply exactly the same procedure that we described in step 3 for long observation periods ($T \geq 30$ s). Then, we can use either the SMPF or the DMPF computed to localize our attention in the IPF spectrum by multiplying $P_{\mathrm{rb}}\prime$ with $\mathrm{N}(\mu_{\mathrm{p}},\sigma^2_{\mathrm{p}})$ that is denoted as $P_{\mathrm{rb}}\prime\prime$. The tentative IPF is the frequency f_i for which the amplitude of the spectrum $P_{\mathrm{rb}}\prime\prime$ is maximum.

Adaptive Estimation Filter and Pulse Recovery: The instantaneous computation described by both ALM and ACM suffers by occasional thermo-regulatory vasodilation and noise despite the effective mechanisms built into both models. This problem has been addressed by building an estimation function that takes into account the current measurement as well as a series of past measurements. This idea is based on the adaptive line enhancement method reported in [3]. In our previous studies, we reported that the current power spectrum of the temperature signal is being computed over the previous 2^9=512 frames by applying the ALM or ACM models. Now, we investigate a frame range from 128 and up to 2,048 frames (2^N frames for $N \in [7, 11]$).

To compute the pulse frequency, first we convolve the current power spectrum computed by either model with a weighted average of the power spectra computed during the previous 60 frames. This is because at the average speed of 30 fps sustained by our system, there is at least one full pulse cycle contained within 60 frames even in extreme physiological scenarios. Then, we compute the historical frequency response (HFR) at a particular frequency. HFR is given as the summation of all the corresponding frequency responses for the spectra, normalized over the total sum of all the frequency responses for all the historical spectra. Finally, we convolve the HFR with the current power spectrum and we then designate as pulse the frequency that corresponds to the highest energy value of the filtered spectrum within the operational frequency band.

References

1. (2004) PowerLab Owner's Manual. ADInstruments Pty Ltd.
2. (2006) SC6000 Owner's Manual. FLIR Inc.
3. Akay M, Akay Y, Welkowitz W, Semmlow J, Kostis J (1992) Application of adaptive filters to noninvasive acoustical detection of coronary occlusions before and after angioplasty. IEEE Trans Biomed Eng 39(2):176–184
4. Aminian K, Thouvenin X, Robert P, Seydoux J, Girardier L (1992) A piezoelectric belt for cardiac pulse and respiration measurements on small mammals. In: Proceedings of the 14th annual international conference of the IEEE engineering in medicine and biology society, 2663–2664
5. Buddharaju P, Pavlidis I, Tsiamyrtzis P, Bazakos M (2007) Physiology-based face recognition in the thermal infrared spectrum. IEEE Trans Pattern Anal Mach Intell 29(4):613–626
6. Chekmenev S, Farag A, Essock E (2007) Thermal imaging of the superficial temporal artery: An arterial pulse recovery model. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, Minneapolis, MN, 1–6
7. Dowdall J, Pavlidis I, Tsiamyrtzis P (2007) Coalitional tracking. Comput Vis Image Underst 106(2-3):205–219
8. Duda R, Hart P, Stork D (2001) Pattern classification. Wiley-Interscience, New York
9. Evans D, McDicken W (2000) Doppler ultrasound: physics, instrumentation and signal processing, 2nd edn. Wiley, Chichester
10. Garbey M, Merla A, Pavlidis I (2004) Estimation of blood flow speed and vessel location from thermal video. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, vol 1. Washington, DC, pp 356–363
11. Garbey M, Sun N, Merla A, Pavlidis I (2007) Contact-free measurement of cardiac pulse based on the analysis of thermal imagery. IEEE Trans Biomed Eng 54(8):1418–1426

12. Guyton A (1991) Textbook of medical physiology, 8th edn. Philadelphia, PA, chap 3
13. Loukogeorgakis S, Dawson R, Phillips N, Martyn C, Greenwald S (2002) Validation of a device to measure arterial pulse wave velocity by a photoplethysmographic method. Physiol Meas 23:581–596
14. Moxham B, Kirsh C, Berkovitz B, Alusi G, Cheeseman T (2002) Interactive head and neck (CD-ROM). Primal Pictures, Castle Hill, NSW
15. Murthy R, Pavlidis I (2006) Noncontact measurement of breathing function. IEEE 25(3):57–67
16. Pavlidis I, Levine J (2001) Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing. In: Proceedings of the 23rd annual international conference of the IEEE engineering in medicine and biology society, vol 3. Istanbul, Turkey, pp 2826–2829
17. Pavlidis I, Dowdall J, Sun N, Puri C, Fei J, Garbey M (2007) Interacting with human physiology. Comput Vis Image Underst 108(1-2):150–170
18. Pinar YA, Govsa F (2006) Anatomy of the superficial temporal artery and its branches: Its importance for surgery. Surg Radiol Anat 28:248–253
19. Sun N, Pavlidis I, Garbey M, Fei J (2006) Harvesting the thermal cardiac pulse signal. In: Medical image computing and computer-assisted intervention-MICCAI 2006, Lecture notes in computer science, vol 4191. Springer, Heidelberg, pp 569–576
20. Bourlai T, Buddharaju P, Pavlidis I, Bass B (Nov 2009) On Enhancing Cardiac Pulse Measurements Through Thermal Imaging. IEEE International Conference on Information Technology and Applications in Biomedicine, Larnaka, Cyprus

Parallel Multispectral Image Segmentation for Computer Aided Thyroid Cytology

Shishir Shah and Edgar Gabriel

Introduction

Cancer continues to remain a major health problem in the United States, with one of two men and one of three women developing cancer in their lifetime. Thyroid nodule is one of many common cancers. Nodules are more common in women and increase in frequency with age and with decreasing iodine intake. It has been estimated that up to 20% of the world population [1,2] and approximately 50% of 60-year-old persons [4] have palpable thyroid nodule or nodules. In the US, up to 7% of the adult population has a palpable thyroid nodule [3]. The clinical spectrum ranges from the incidental, asymptotic, small, solitary nodule, in which the exclusion of cancer is a major concern, to the large, partly intrathoracic nodule that causes pressure symptoms, for which treatment is warranted regardless of cause [4,5] . The most common cytologic diagnoses of thyroid nodules are colloid nodules, cysts, thyroiditis, follicular neoplasm, and thyroid carcinomas. Colloid nodules are the most common and do not have an increased risk of malignancy, therefore, the choice of management is conservative. Follicular neoplasm includes follicular adenoma and follicular carcinoma, which cannot be distinguished visually from each other based on cytology and the management remains controversial [2, 4, 5]. Thyroid carcinoma occurs in roughly 10% of all palpable nodules and the management is surgical removal [41].

Recent advances in clinical oncology have significantly increased the survival rate for most cancer patients. Early detection of the disease has been partly responsible for improved outcomes. Cytological and histological assessment has played an integral and important role in cancer diagnosis and prognosis. It is critical for the detection and characterization of excised tissue and cells from nodules and to identify its malignancy level. Among early screening methods, fine needle aspiration (FNA) has been widely accepted as a first-line screening test in patients with thyroid nodule. The major role of thyroid fine needle aspiration cytology (FNAC) is to distinguish thyroid nodules that require surgical removal due to higher

S. Shah (✉) and E. Gabriel
Department of Computer Science, University of Houston, Houston, TX 77204, USA
e-mail: sshah@central.uh.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_7,

risk of malignancy from those benign lesions that can be managed conservatively. According to guidelines from American Association of Clinical Endocrinologists, FNA is believed to be the most effective method available for distinguishing between benign and malignant thyroid nodules [6]. Its utilization has increased in recent years partially because of increased diagnostic accuracy and wide availability of image-guided techniques, which allow smaller nodules to be detected and aspirated. In centers with experience in FNA, the use of this technique has been estimated to reduce the number of thyroidectomies by approximately 50%, to roughly double the surgical confirmation of carcinoma, and to reduce the overall cost of medical care by 25% [40], as compared with surgery performed on the basis of clinical findings alone.

In general, cytological diagnoses of thyroid disorders are divided into following categories: unsatisfactory/nondiagnostic, benign/negative for malignancy, indeterminate for malignancy, suspicious for malignancy, and positive for malignancy [7]. This in turn has exerted an increased impact on histological assessment in conjunction with early screening methods, moving it from an ancillary diagnostic tool to a stand-alone diagnostic procedure. FNAC is progressively replacing other methods such as radionucleotide scanning, because of more accurate diagnosis, expediency in obtaining treatment, overall cost of patient care, and smaller biopsy samples. As such there has been a concomitant increase in the demand for tools for assessment of FNAC.

Challenges for FNAC Diagnosis

The cytodiagnosis of thyroid nodules by FNA is complex for the following reasons [8]:

- Overlap of cytological patterns between neoplastic and nonneoplastic lesions
- Overlap of cytological features between neoplasm's
- Coexistence of nonneoplastic and neoplastic processes and multiple malignancies in the same gland

Currently, the "gold standard" for diagnosis of follicular adenoma and follicular carcinoma are histology. In many cases, the distinction between them can be difficult even on the postoperative histologic specimen [9]. Indeterminate or follicular lesions are thyroid lesions described as having cytologic evidence that may be compatible with malignancy but are not amenable to easy diagnosis. The American Thyroid Association clearly presents the controversies surrounding this problem in the laboratory medicine practice guidelines [10]. Majority of cytopathologists agree that cytologic differential diagnosis between these pathologies is often difficult because of the lack of visually distinctive elements and overlapping morphology [11]. The recommendation for managing patients with nodules in this category is surgery. However, only about 30% of the cases are proved to be malignant on histology, remaining being benign which would be better managed by clinical follow-up

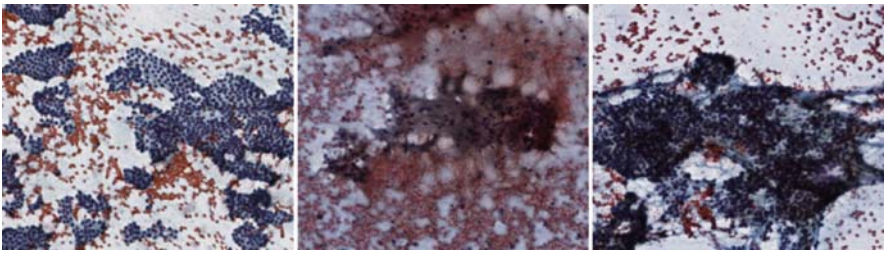


Fig. 1 Follicular lesions with varying cellular patterns and morphology

instead of surgery. Currently, there are no reliable methods to differentiate follicular adenoma from follicular carcinoma before surgery. Figure 1 shows example of cytology images with varying cellular patterns and morphology for indeterminate cases in follicular thyroid nodule.

Multispectral Imaging

Multispectral imaging combines two technologies that have independently evolved for the last several decades, namely spectroscopy and imaging. Systems that leverage the two technologies simultaneously have been used in the fields of astronomy [12], remote sensing [13], and chemical compound analysis [14] to identify the composition and characteristics of celestial, terrestrial, and atmospheric elements. In recent years, spectral imaging has also been used in biomedical applications such as spectral karyotyping [15], general cell visualization [16], cell trafficking of variously colored fluorescent proteins [17], and differentiation of pathologies [18]. Spectral imaging systems developed to facilitate various biomedical applications have typically been coupled with traditional microscopes to enable spectral microscopy. Spectral microscopes provide the combined benefits of spectroscopy and imaging microscopy, resulting in the ability to acquire spectral images of microscopic specimen.

More recently, multispectral microscopes capable of acquiring spectral images under transmitted illumination have also been used to digitize and analyze cell smears [19,20]. Spectral imaging allows for the simultaneous measurement of spectral and spatial information of a sample such that the measurement of the spectral response at any pixel of a two-dimensional image is possible. Figure 2 shows a single field-of-view of a thyroid FNA cell smear. Shown are four spectral bands acquired using an upright microscope attached to a monochromator for narrowband specimen illumination. Studies have shown that biological tissue exhibits unique spectra in transmission. By exploring the spectral differences in cellular pathology, many chemical and physical characteristics not revealed under traditional imaging systems can be realized and used to improve the analysis efforts.

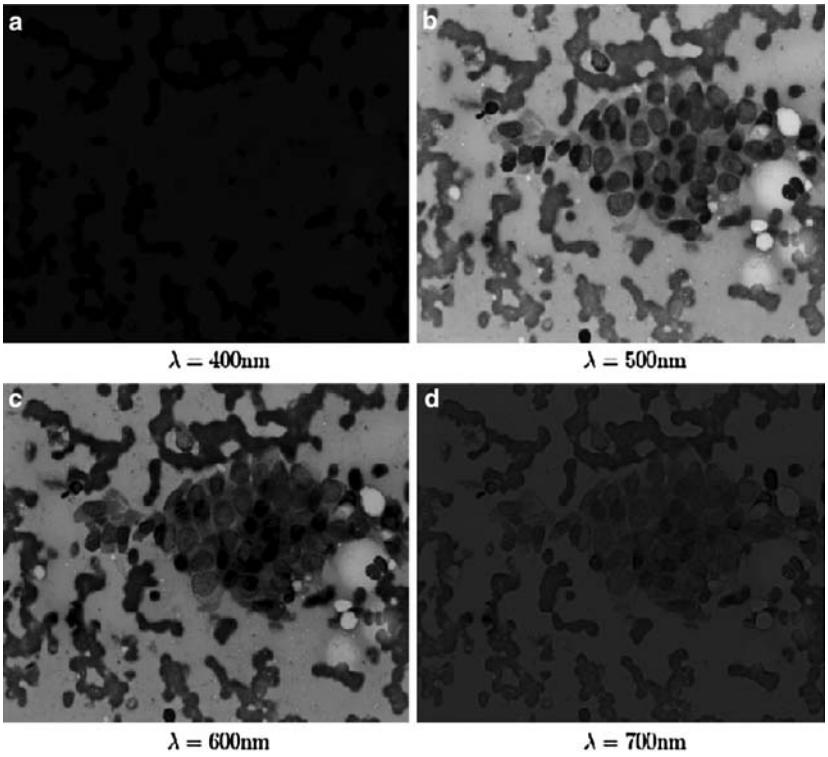


Fig. 2 Four bands for a multispectral image acquired for a single field-of-view of a thyroid fine needle aspirated cell smear

Computer-Aided Analysis Needs for Thyroid FNAC

Determination of a diagnosis based on FNA cytology is driven primarily by the expertise and experience of the reviewing cytopathologist. This process is inherently subjective and imprecise. Moreover it is slow and laborious, and yields only qualitative results that are difficult to reproduce, and do not reflect the accurate differential diagnosis that is so desirable [21]. It is clear that quantitative measurements are warranted to allow improved and reliable diagnosis. Thus, computer-assisted analysis is slowly gaining popularity as a more accurate, unbiased, and expedient method of quantitation. Consequently, several commercial entities have developed a variety of hardware and software solutions to the more general problem of histological examination. However, widespread adoption and application of the technology remains unfulfilled because of high instrumentation costs and a lack of methods for automated quantitative analysis.

We present a multispectral microscopy system capable of acquiring spectral images under transmitted illumination and algorithms leading to automated analysis of thyroid FNAC. Several key steps have to be performed to automate this task, including image acquisition, cell segmentation, feature extraction, and

classification. While each of the steps is important to the overall success of the system, image segmentation is known to be a critical problem that impacts the accuracy of any automated image analysis system. Hence, in this chapter, we specifically discuss the challenges and present an approach to cell segmentation in multispectral images. Along with algorithmic challenges, another limitation for the deployment of an automated solution for the analysis of FNAC images is the overall size of a multispectral image and the resulting processing time for the image segmentation. Multispectral FNAC images often contain tens or even hundreds of spectral channels, increasing the data volume to be analyzed to tens of Gigabytes of raw data. Even in case the PC utilized for the segmentation of such an image has enough main memory to perform the analysis, the procedure might take multiple hours, making it unfeasible to utilize it in an everyday production environment with near real-time constraints.

However, recent developments in processor architectures, most notably multicore processors and compute accelerators such as general purpose graphical processing units (GPGPUs), offer the potential to speed up the analysis significantly. Multicore processors contain several independent "processors," each of which can execute either a separate sequential application or different processes/threads of a single parallel application. In order to reduce the execution time of an image segmentation code for a large multispectral image, the latter approach has to be considered. However, the application needs to be restructured and partially rewritten in order to take advantage of the increased computational power of such a multicore processor. Most typically, the application has to be broken up into different pieces which can be executed independently, e.g., on the data level, such that one process/thread works on the first part of the image, while the second process/thread works on the second part, etc. An important question arising for the parallelization of an application is whether to utilize a multithreaded or a multiprocess approach. Within this chapter, we will compare a multithreaded parallelization approach using OpenMP [22] to a multiprocess parallelization strategy using the message passing interface (MPI) [23]. We compare both parallel versions conceptually as well as their performance on a state-of-the-art PC containing two quad-core AMD Opteron processors (eight cores total) with 8 GB of main memory.

Methods

Spectral Imaging

The acquisition of spectral images of a FNAC smear is typically performed using an imaging system comprising of a standard light microscope coupled to a spectral dispersion component that separates the light into its spectral components. The microscope itself is coupled to a two-dimensional (2D) optical detector such as a CCD camera, or to an array of photomultiplier tubes (PMT). In our case, we have used a grating based spectral light source coupled to a standard optical microscope allowing 2D image acquisition using a high resolution CCD camera.

Specifically, we use a quarter-meter class, Czerny–Turner type monochromator from PTI (http://www.pti-nj.com) that provides a tunable light emission spectrum at 2 nm resolution. We currently utilize a wavelength range from 400 to 700 nm. The monochromator is connected to an Olympus (http://www.olympus.com) BX51 upright optical microscope such that the light output from the monochromator feeds into the transmitted light path of the microscope. An Olympus UPlanApo 40X NA 0.9 is used for imaging. The Photometrics (http://www.roperscientific.com) SenSys^{TM} CCD camera having 768 × 512 pixels (9 × 9 μm) at 8-bit digitization is used, which provides for high-resolution low light image acquisition. This camera is used for acquisition of spectral images.

To image each sample, the illumination from the monochromator is adjusted by achieving Kohler illumination for uniform excitation of the specimen. The condenser, aperture diaphragm, and the field stop are kept constant during measurements. To calibrate the variations due to illumination effects and varying quantum efficiency of CCD cameras at different wavelengths, we normalize the camera exposure for each wavelength while avoiding saturation. The underlying assumption is that the incident light should be uniform across all wavelengths. Since measurement of incident light is proportional to average background intensity, the problem of calibration is posed as computing exposures for each wavelength such that the resultant exposure values across the spectra generate uniform average background intensity. To achieve this objective, we pose the problem as one of equalizing the image average optical density (AOD) across all the wavelength-exposure pairs. Wavelength-exposure pair images are the set of varying exposure images for each wavelength. A least square error solution is chosen to compute the final exposure values across the entire wavelength range [24]. Following calibration, automatic focusing is performed at the central wavelength of 550 nm to minimize the chromatic aberration at all wavelengths. Spectral image set is acquired by stepping the monochromator from 400 to 700 nm illumination in 10 nm increments, resulting in a total of 31 images for each field-of-view.

Considering an average size of the smear to be 1.0 cm × 1.0 cm under 40× magnification and a resolution of 0.25 μm/pixel, and assuming furthermore 5–31 spectral channels for an image, the data cube to be analyzed would be approximately 8–50 GB in size. This creates difficulties in analyzing the entire data set on a standard desktop, especially when we take into account, that the code required to perform the according analysis will have to store typically a multiple of the raw data size of the overall image for intermediate processing.

Image Segmentation

Image segmentation is probably the most widely researched topic in image analysis and many attempts have been made to develop algorithms for segmentation of biomedical images. This is a critical problem since it forms the first step in identifying cells and tissue structures relevant for subsequent analysis. It is also the

most challenging task due to the variabilities present in the images to be segmented. The most common approach for cytological image segmentation has been thresholding with the purpose of separating cells from the background. The problem of overlapping and touching cells is typically solved using a second step based on mathematical morphology [25] or watershed algorithms [26]. In some cases, learning algorithms coupled with clustering techniques have been used where pixels are assigned to either "cell" or "noncell" class [27, 28]. Other mathematical formulations ranging from active contours, Fourier and Hough transforms, neural networks, and others [29–32] have also been developed. An overview of segmentation techniques, specifically for cell segmentation can be found in [33]. In general, there is no single segmentation algorithm that can be considered good for all images, nor are all algorithms equally good for a particular image [34]. In fact, the utility of each algorithm is limited by their narrow, highly specific characteristics that make them applicable for certain kinds of images.

In this work, we discuss the segmentation of cells in spectral images. We leverage a pattern theoretic approach of feature extraction for each pixel followed by clustering to achieve separation of cell pixels from those belonging to background. Finally, the pixels clustered as belonging to cells are refined using morphological operations. We use a bank of Gabor filters to extract a measure of texture at each pixel followed by k-means clustering to group pixels belonging to the same class. Specifically, we generate a bank of self-similar filters through appropriate dilations and translations of the basis Gabor function as defined by Manjunath and Ma [35]. We use three scales and four orientations resulting in a total of 12 filters in the bank. To efficiently compute a measure of texture for each pixel in the multispectral image, we generate an average image from the multispectral stack. For each pixel in the average image, the magnitude response of each filter, the mean, and standard deviation is computed and stored as a feature vector. In addition, the absorption is measured for each pixel in each channel of the multispectral image, given as:

$$A = \log(1/T) \tag{1}$$

where A is the absorption and T is the transmission measured as the ratio of the transmitted light to the incident light. Hence, we generate a feature vector for each pixel belonging to a spectral image of dimension number_of_channels $+14$. The extracted features are clustered using the standard k-means algorithm [27] that results in effective grouping of pixels belonging to the thyroid cells and partitioning of the image. Morphological filtering is performed to remove small clusters as the final step. Figure 3 (left) shows an example of a thyroid cell smear image at 520 nm (1 of 31 channels) and the result of clustering (right) with the cluster of pixels detected as thyroid cells and refined by morphological smoothing overlayed in black.

To give an indication of the challenges that the image segmentation step poses from the computational perspective, and a motivation for the subsequent parallelization step, consider the memory consumption for code performing the operation described above. In order to determine the filter response using Fast Fourier Transform operations (FFT), the code will have to allocate typically three arrays of the

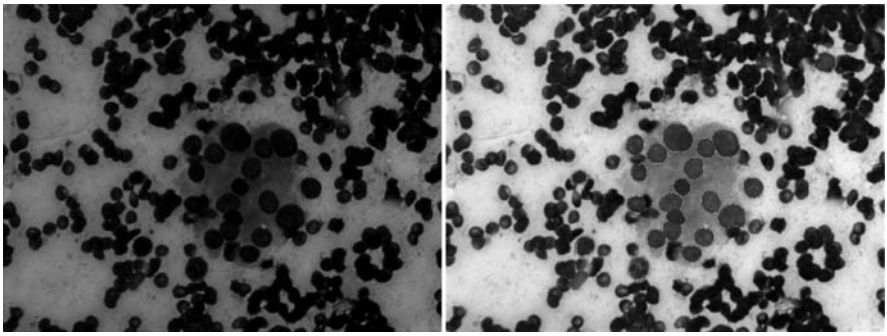


Fig. 3 One image from a multispectral cell smear stack (*left*) and the corresponding result of segmentation highlighted by bounding contours around the detected cells of interest (*right*)

dimension of the overall image, with each data point being a double complex value. Furthermore, the feature vector used for the subsequent clustering step consists of (number_of_channels +14) dimensions for each pixel. For an image consisting of 1,024 × 1,024 pixels and 31 spectral channels, the memory consumption of these arrays will be 424 MB. For an image twice that size, i.e. 2,048 × 2,048 pixels, the memory consumption exceeds 1.65 GB, and if the number of pixels is doubled once more in each dimension, the code has to allocate over 6.6 GB for the feature vector and the FFT operations. Note, that realistically the memory consumption will be even higher due to temporary arrays in the FFT library and other parts of the application. Additionally to the memory constraints, the computational constraints also increase with the size of the image, since most of the compute intensive operations do not scale linearly with the number of pixels, but have higher orders of computation complexities. For example, the computational complexity of an FFT operation is O(N log N) for an image consisting of N pixels, and O(kN) for k-means clustering having k clusters. Therefore, the time required to perform the analysis will increase dramatically with the image size, making the deployment of a CAD solution with near real-time requirements highly challenging.

Parallelization

In order to fully exploit the capabilities of modern multicore processors when analyzing large multispectral images, two different parallel versions of the image segmentation code have been created. The first version is based on the MPI [23], which uses multiprocess approach, and a second version based on OpenMP [22], the de-facto standard for shared memory programming in scientific computing. The sequential code used as the basis for both parallel versions is implemented in C, using the FFTW library [36] for the Fast-Fourier Transform (FFT) operations. The parallel versions described here achieve the same results with the same accuracy as their sequential counterparts.

As with most MPI applications, the parallelization strategy used in this version of the code relies on data decomposition: each process holds one part of the overall image, and all processes execute the same code on different data items. FFTW version 2.1.5 supports MPI parallel FFT operations. However, the library mandates one-dimensional data decomposition, i.e., each process holds a certain number of rows of the overall image. Furthermore, the size of the image passed to the FFTW routines have to be padded by the number of rows/columns of the Gabor filters, resulting in a slightly uneven distribution of rows for the unpadded image across the processes.

The most challenging part of the MPI version was the parallelization of the k-means clustering routine. Assuming that the number of clusters is small compared to the number of pixels, the cluster specific information has been replicated on all processes. Each process determines locally for each pixel in its domain the cluster whose center is closest to each pixel and assigns that pixels to the appropriate cluster. The code also determines the number of pixels assigned to each cluster and the weight of each cluster. Following these local calculations are three global reduction operations, which determine the overall number of pixels assigned to each cluster across all processes, the overall weight of each cluster, and the global error, defined as the sum of the squared distance of each pixel to the center of the closest cluster. Using these global values, each process can independently determine the new center of each cluster for the next iteration of the algorithm. This iterative procedure is terminated when the error between two iterations is smaller than a predefined threshold.

The code then performs a smoothing operation by comparing the cluster assigned to a particular pixel with the clusters assigned to its neighboring pixels. In case the majority of the neighboring pixels belong to a cluster, which is different than the current cluster assigned to the analyzed pixel, that pixel is reassigned to the cluster of surrounding pixels. In order to implement this procedure in parallel, each process will need access to the cluster information of pixels at the boundaries of its computational domain, which is located on another process. In the MPI version access to information owned by another process is realized by introducing "ghostcells," i.e., introducing copies of information (in this case which cluster does a pixel belong to) owned by another process. The size of the ghostcells is determined by the number of neighboring cells analyzed.

The OpenMP parallel version of the code also follows a data decomposition approach. FFTW version 1.2.5 also supports OpenMP type thread level parallelism, which eased the parallelization of the filtering and convolution, as it only required an additional parameter (number of threads) to be passed to the functions. However, the initial version of this code section did not perform well. Profiling the application with the performance analysis tool TAU [37] revealed that OpenMP directives inserted for a loop performing the padding of the image degraded the performance due to a large number of cache misses between the different threads. Thus, the performance of this code section could be improved significantly by not parallelizing it.

The k-means clustering revealed a series of other challenges. Following a similar approach as in the MPI version, OpenMP directives were inserted to parallelize the access to the individual pixels. Determining the number of pixels assigned to

a cluster and the weight of each cluster across all threads posed, however, some problems, since OpenMP does not support reductions over arrays in C in the current specification. Different alternatives have been explored to overcome this limitation including introducing critical regions around the clustering arrays, atomic updates, and locking only a particular value of the clustering data. The version leading to the best performance privatized the clustering arrays and performed global updates by executing element-wise reduction operations. For more details on both parallelization approaches, please refer to [38].

Results

In the following we evaluate the performance of the MPI as well as the OpenMP version of the code, and compare them to each other. The platform chosen resembles a typical high-end PC, having two 2.2 GHz quad-core AMD Opteron processors (eight cores total), with 8 GB of main memory. The PC runs a wide-spread Linux distribution (OpenSuSE 11.0), using the gcc compiler suite in version 4.3 and Open MPI [39] version 1.2.6. Tests have been executed for two images, namely an image consisting of 1,024 × 1,024 pixels with 21 spectral channels, and an image consisting of 2,048 × 2,048 pixels with 21 spectral channels and an overall image size of 84 MB. The parallel codes have been executed using one, two, four, and eight threads/processes, respectively, in order to determine the parallel scaling of the according version. Each test presented in the following has been executed three times, and the minimum over the three runs has been used for generating the graphs.

Figure 4 shows the overall execution time of the MPI as well as the OpenMP version of the code for both images. Both parallel code versions scale very well,

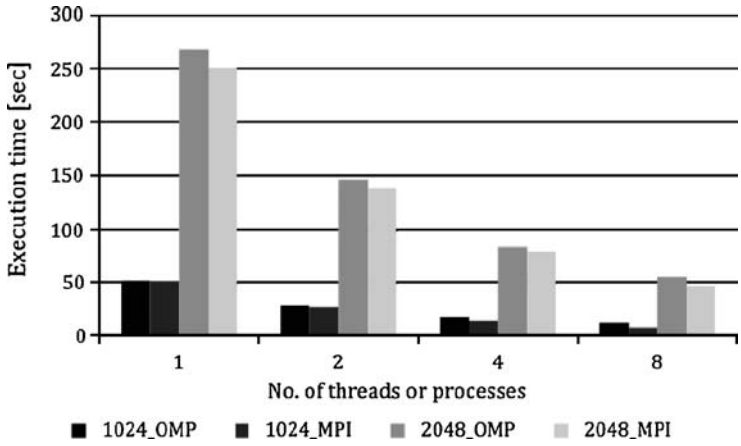


Fig. 4 Overall execution time of the MPI and OpenMP versions of the code for both images and various numbers of processors, respectively, threads

i.e., increasing the number of compute cores utilized by the application reduces the overall time required to analyze the according image. As an example, the execution time of the larger image is decreased from 267 s using one thread to 54 s using eight threads for the OpenMP version, a performance improvement by a factor of 4.9. For the very same image, the MPI version reduces the execution time from 250 s when using one process to ∼47 s when using eight processes, a speedup of 5.3. The MPI version of the code slightly outperforms the OpenMP version of the code. In order to understand which operations are responsible for this behavior, a detailed performance study of the most time-consuming operations of this application has been performed. Figures 5 and 6 show the execution time of the convolution operation and of the k-means clustering algorithm for both parallel code versions.

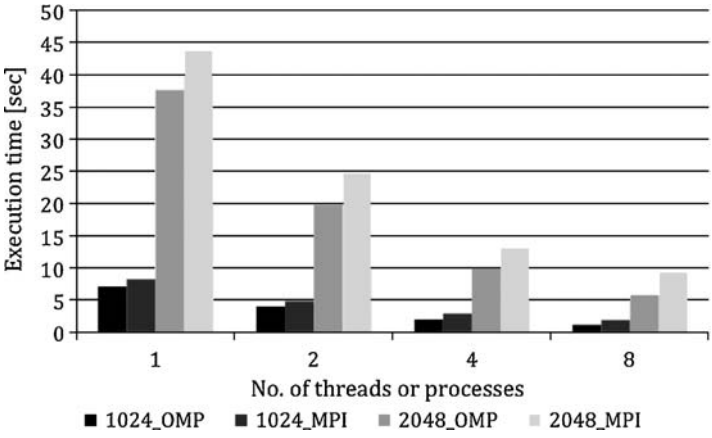


Fig. 5 Execution times of the convolution operation for both parallel versions

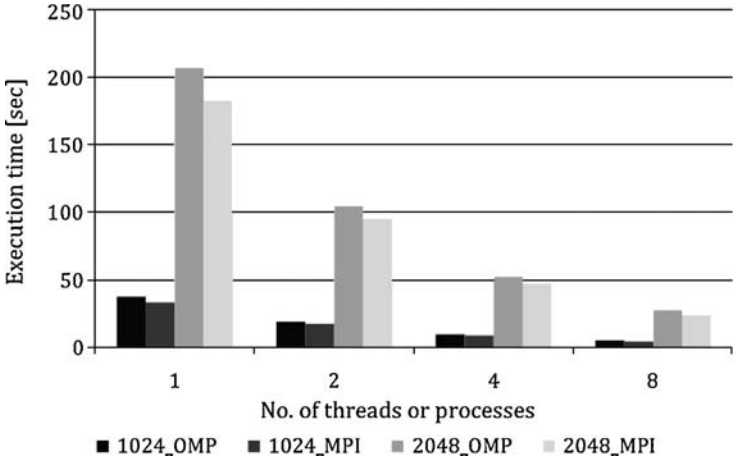


Fig. 6 Execution times of the k-means clustering for both parallel versions

The results indicate that the convolution operation, which contains a large number of FFT operations, is faster when using the multithreaded code version based on OpenMP. The reason for this is that MPI parallel FFT operations requires all-to-all type data exchange between the processes, which is considered expensive even for shared memory architectures such as a multicore PCs. These all-to-all communication operations are not required in the multithreaded parallelization, since all threads have access to exactly the same memory locations.

The situation is reversed for the k-means clustering algorithm. In this case, the separation of the address space for the different MPI processes helps in avoiding synchronization between the threads due to concurrent access to the same memory location and unnecessary false sharing of cache lines occurring in the multithreaded version. False sharing occurs when different compute cores have separate caches which can hold, under certain circumstances, copies of the same data items. As soon as a thread modifies a data item in the cache, the copy of the very same data item in the cache of the other compute cores has to be invalidated (or updated), since it is considered to be outdated. However, the cache hardware operates on full cache lines, which are typically 64 bytes long. This means, that the hardware invalidates an entire cache line containing multiple data item in case one of them has been modified. If multiple threads modify different data items which are residing, however, in the same cache line, a large number of cache invalidations are occurring, which are unnecessary considering that threads do not access the same data item. This phenomenon is known as false sharing, and it affects mostly multithreaded applications. Overall, the performance benefit of the MPI version in the k-means clustering outperforms the performance benefit of the OpenMP version in the convolution operation, and leads thus to an overall better performance of the MPI version.

Conclusions

Within this chapter, we have discussed the need for and the current status of automated solutions in cancer detection, focusing especially on FNAC for thyroid nodules. A multispectral microscopy system is discussed and processing challenges presented that pose a bottleneck for the development of a clinically viable CAD solution. The resulting images required to be analyzed are often tremendous in size, a circumstance further emphasized by multispectral images, which allow for the simultaneous measurement of spectral and spatial information of a sample. On the other hand, today's computing systems based on new multicore processor technologies offer the opportunity to solve enormous, compute-intensive problems on a regular desktop PC. However, in order to utilize the computational power of these systems, programmers have to extend their application by explicitly denoting the parallelism available in their application. The focus of this chapter is on how to bridge the gap between image processing and modern computer architecture, thereby meeting the demands imposed by this highly challenging class of applications.

For this, we evaluated two alternative parallelization strategies for an image segmentation code. An OpenMP version which utilizes multiple threads and an MPI version based on multiple, concurrent processes has been developed. Experiments on a state-of-the-art multicore PC shows the scalability of both solutions, offering significant performance advantages compared to the sequential version. Hence, executing the image segmentation code on large images is now possible in a short time frame. This represents a significant step towards a fully automated CAD solution.

References

1. Hegedus L (2004) The thyroid nodule. New Eng J Med 351:1764–1771
2. Hegedus L, Bonnema SJ, Bennedbaek FN (2003) Management of simple nodular goiter: current status and future perspective. Endocr Rev 24:102–132
3. Singer PA, Cooper DS et al (1996) Treatment guidelines for patients with thyroid nodules and well differentiated thyroid cancer. Arch Intern Med 156:2165–2172
4. Bennedbaek FN, Perrild H, Hegedus L (1999) Diagnosis and treatment of solitary thyroid nodule: results of a European survey. Clin Endocrinol (Oxf) 50:357–363
5. Bennedbaek FN, Perrild H, Hegedus L (2000) Diagnosis and treatment of solitary thyroid nodule: results of a North American survey. J Clin Endocrinol Metab 85:2493–2498
6. Feldman JA (1982) Dynamic connections in neural networks. Biol Cybern 46:27–39
7. Papanicolaou Society of Cytopathology Task Force on Standard of Practice (1996) Guidelines of the papanicolaou society of cytopathology for the examination of fine-needle aspitation specimens from thyroid nodules. Mod Pathol 9:710–715
8. McCall A, Jarosz H, Lawrence AM, Paloyan E (1986) The incidence of thyroid carcinoma in solitary cold nodules and multinodular goiter. Surgery 100:1128–1132
9. Rosai J, Carangiu ML, DeLillis RA (1990) Atlas of tumor pathology. Armed Forces Institute of Pathology, Washington
10. The American Thyroid Association (2003) Consensus guidelines for thyroid testing in the new millennium
11. Gutman PD, Henry M (1998) Fine needle aspiration cytology of the thyroid. Clin Lab Med 18:461–482
12. Kim E, Choi Y, Yang HS, Kang M, Jeong S, Yang S, Kim ED, Lee JH, Rasheed AA, Nasir HM, Rosdi MR (2003) A high-resolution multi-spectral imaging system for small satellites. Acta Astronaut 52(9):813–818
13. Kerekes J, Baum J (2002) Spectral imaging system analytical model for subpixel object detection. IEEE Trans Geosci Remote Sens 40(5):1088–1101
14. Kulcke A, Gurschler C, Spock G, Leitner R, Kraft M (2003) On-line classification of synthetic polymers using near infrared spectral imaging. J Near Infrared Spectrosc 11(1):71–81
15. Schrock E, Manoir SD, Veldman T, Schoell B, Weinberg J, Ferguson-Smith M, Ning Y, Ledbetter D, Bar-Am I, Soenksen D, Sarini Y, Reid T (1996) Multicolor spectral karyotyping of human chromosomes. Science 273:494–497
16. Maiti D, Sennoune S, Martinez-Zaguilan R (2003) Proton gradients in human breast cancer cells determined by confocal, multiphoton, and spectral imaging microscopy. FASEB J 17(4):A467
17. Haraguchi T et al (2002) Spectral imaging fluorescence spectroscopy. Genes Cells 7(9): 881–887
18. Roy K, Bottrill I, Ingrams D, Pankratov M, Rebeiz E, Woo P, Shapshay S, Kabani S, Manoharan R (1995) Diagnostic fluorescence spectroscopy of oral mucosa. Proc SPIE 2395:135–142
19. Feng C, Shuzhen C, Libo Z (2007) New abnormal cervical cell detection method of multispectral pap smears. Wuhan Univ J Nat Sci 12:476–480

20. Shah S, Schwartz MR, Mody DR, Scheiber-Pacht M, Amrikachi M (2008) The role of multispectral microscopy in differentiating benign and malignant thyroid nodules: a pilot study of 24 cases. In: Proceedings of the annual meeting of the United States and Canadian academy of pathology
21. Suen KC (2002) Fine-needle aspiration biopsy of the thyroid. Can Med Assoc J 167:491–495
22. Chapman B, Jost G, Pas R. v. d. (2008) Using OpenMP: portable shared memory parallel programming. MIT, Cambridge, MA
23. Message Passing Interface Forum (2008) MPI: A Message Passing Interface Standard version 2.1.
24. Shah S, Thigpen J, Merchant F, Castleman K (2006) Photometric calibration for automated multispectral imaging of biological samples. In: Proceedings of 1st workshop on microscopic image analysis with applications in biology (in conjunction with MICCAI, Copenhagen), 27–33
25. Serra J (1982) Image analysis and mathematical morphology. Academic, London
26. Vincent L, Soille P (1991) Watersheds in digital spaces: An efficient algorithm based on immersion simulations. IEEE Trans Pattern Anal Mach Intell 13:583–598
27. Hartigan JA, Wong MA (1979) A K-means clustering algorithm. Appl Stat 28:100–108
28. Faugeras OD, Pratt WK (1980) Decorrelation methods of texture feature extraction. IEEE Trans Pattern Anal Mach Intell 2:323–332
29. Bamford P, Lovell B (1998) Unsupervised cell nucleus segmentation with active contours. Signal Process 71:203–213
30. Kurugollu F, Sankur B (1998) Color cell image segmentation using pyramidal constraint satisfaction neural network. IAPR Workshop Mach Vis Appl 85–88
31. Lu J, Li S, Yang J, Liu L, He A, Liu K (1998) Lung cancer cell recognition based on multiple color space. Proc SPIE 3522:378–386
32. Pardo XM, Cabello D (2000) Biomedical active segmentation guided by edge saliency. Pattern Recognit Lett 21:559–572
33. Ablameyko S, Nedzved A, Lagunovsky D, Patsko O, Kirillov V (2001) Cell image segmentation: review of approaches. Proc ICPR 2:26–34
34. Pal NR, Pal SK (1993) A review on image segmentation techniques. Pattern Recognit 26: 1277–1294
35. Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell (PAMI – Special issue on Digital Libraries) 18:837–842
36. Frigo M, Johnson SG (2005) The design and implementation of FFTW3. Proc IEEE 93: 216–231
37. Malony AD, Shende S, Bell R, Li K, Li L, Trebon N (2004) Advances in the TAU performance system. Perform Anal Grid Comput 129–144
38. Gabriel E, Venkatesan V, Shah S (2008) Towards high performance cell segmentation in multispectral fine needle aspiration cytology of thyroid lesions. In: 11th international conference on medical image computing and computer assisted intervention, workshop on high-performance medical image computing and computer aided intervention, New York, NY, USA
39. Gabriel E, Fagg GE, Bosilca G, Angskun T, Dongarra JJ, Squyres JM, Sahay V, Kambadur P, Barrett B, Lumsdaine A, Castain RH, Daniel DJ, Graham RL, Woodall TS (2004) Open MPI: goals, concept, and design of a next generation mpi implementation. In: Proceedings, 11th European PVM/MPI Users' Group Meeting, Budapest, Hungary
40. Mazzaferi EL (1992) Thyroid cancer in thyroid nodules: finding a needle in the haystack. Am J Med 93:359–362
41. Mazzaferi EL (1993) Management of a solitary thyroid nodule. N Engl J Med 328:553–559

Part III
Image Driven Intervention and Robotics

Computer-Assisted Digestive Surgery

Luc Soler, Stéphane Nicolau, Alexandre Hostettler, Jean-Baptiste Fasquel, Vincent Agnus, Arnaud Charnoz, Johan Moreau, Bernard Dallemagne, Didier Mutter, and Jacques Marescaux

Introduction

Introducing an optical device into the abdomen of a patient so as to carry out the surgical procedure via a miniaturized camera represented the major change the surgical world experienced during the twentieth century: the "minimally invasive" surgery era was born. This revolution is about to experience a new twist linked to the appearance of a new original technique called Natural Orifice Transluminal Endoscopic Surgery (NOTES) that could replace traditional laparoscopic surgery for a large set of procedures. By replacing the rigid optic that is introduced through the skin by a flexible optic that is introduced through a natural orifice such as stomach, vagina or colon, this new technique should eliminate all visible incisions.

If the benefits for patients have clearly been proved for laparoscopic surgery, and whatever the result for NOTES, such minimally invasive techniques bring up new difficulties for surgeons, thus reducing their gesture capacity. The first difficulty is the loss of several senses such as the sense of touch and a modification of the force feedback feeling. In NOTES, this loss is greatly amplified due to the length of instruments making it difficult to feel a contact between an instrument and an organ. This lack of force feedback is also featured by current robotic systems, such as the Da Vinci robot from the Intuitive Surgical Company, currently the most used surgical robot worldwide. The use of stereoscopic vision, however, allowed to lessen that perception limit, compensating it by a 3D view of the operative scene filmed by two cameras. But this technique will be difficult to implement for transluminal endoscopic surgery, since it requires the extreme miniaturization of cameras while maintaining a high image resolution. Another solution consists in using virtual reality and augmented reality. Indeed, virtual reality allows to provide a preoperative 3D view of patients, operated from their medical image (CT scan or MRI). This virtual copy of patients can then be used in a preoperative simulator, what provides a realistic 3D view of patients.

L. Soler (✉), S. Nicolau, A. Hostettler, J.-B. Fasquel, V. Agnus, A. Charnoz, J. Moreau, B. Dallemagne, D. Mutter, and J. Marescaux
IRCAD/EITS, 1 place de l'hôpital, 67091 Strasbourg Cedex, France
e-mail: luc.soler@ircad.fr

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_8,

Thanks to an improved preoperative knowledge of each patient's internal anatomy, today practitioners can thus establish an improved diagnosis and a better planning of the best suited therapy for a given case. Therefore, 3D modeling of a patient is generally used for diagnosis support or surgical planning tools. The other use is patient follow-up over time, easing visualization of therapy efficiency. However, surgical simulation still remains today limited to virtual models, without really exploiting medical data of patients. Thus, the simulation of an intervention on a virtual patient reconstructed from his/her medical image represents a major research field, which would allow to reduce medical inaccuracies, thanks to an optimized preoperative training. However, even if the virtual world is a numerical copy of the reality, such preoperative use of virtual reality limits the patient to a virtual world.

In order to overcome this limit, a solution consists in combining information from the virtual world with information from the real world. In medicine, this concept will be translated into two major axes: computer-assisted guiding systems that use real information to control the virtual world, and Augmented Reality systems that superimpose virtual information onto view of the real world. These two axes make up for the lack of sense of touch, thanks to an improved visualized image augmented with virtual information. This computer assistance will therefore allow to find tumors or vessels not by locating them, thanks to touch, but simply by visualizing them.

These techniques can also be used in combination with a robot. Indeed, a robot could replace the human gesture, thanks to an automatic guidance performed by the computer-assisted guiding systems and augmented reality technologies. However, such control remains complex to develop due to the fact that the preoperative models do not contain all required information on the patient. Indeed, all organs and tissues can currently not be modeled preoperatively from CT or MRI. It is typically the case of connective tissue linking organs which are not visible in the medical image. To avoid such a problem, a future solution should consist in combining augmented reality and real-time servoing. Such a link will lead to a real automatic pilot control by preoperative surgical planning automatically adapted to the real patient thanks to a real-time analysis of organ movement and reaction. We propose here to present the state of our work to reach this future objective representing one possible future of surgery.

Virtual Reality for Preoperative Surgical Planning

The first advantage of diagnosis and surgical planning support is to provide a fast, efficient, and easy to implement vision of patient anatomy. Any software meeting these needs will therefore allow to read images realized in clinical routine in their standard format: DICOM. Moreover, that kind of software will have to allow an immediate rendering of at least two kinds: a 2D view of image slices and a 3D view. Several types of 3D volume rendering are being used in routine: direct volume rendering, MIP rendering (Maximal Intensity Projection) and ray casting rendering for

virtual navigation, such as colonoscopy or fibroscopy. There are currently many available software applications that have got to be paid for on visualization consoles of radiology departments, or that can be freely downloaded from the internet. Osirix [1] is the most known and most used software for radiologists. Though it is very comprehensive, it presents two drawbacks: it only works on MacOS and its user interface is not particularly intuitive for surgeons, since it is too close to the software of radiology post-processing consoles. Free or not, we have noticed that these software applications are generally little used by surgeons due to their complexity: the user interface has too many options which are difficult to understand and it requires sometimes a long training of the software. To overcome this recurring drawback, we have developed a free software, VR-Render [2] ©*IRCAD2008*, which has the advantage of being very easy to use with a reduced training time. One of the major benefits for surgeons is direct volume rendering which is automatically computed by the software from the CT or MRI slices of the DICOM image. This technique, available on all current imaging systems (MRI or CT-scan), can be sufficient for a good 3D visualization of anatomical and pathological structures. It consists in replacing the standard slice view by the visualization of all slices together in 3D. In order to see internal structures, the initial grey level is replaced by an associated color and transparency. This transparency gives the feeling of seeing delineated organs which are not delineated in reality. To switch to that mode, the user just uses a simple key and selects automatically computed 3D renderings from a very explicit list. That volume can also be cut along the three main axes (axial, frontal, or sagittal) or with an oblique mouse-controlled plane. In clinical routine, direct volume rendering will be able to offer great preoperative interest. This is the case for all malformation pathologies, in particular vascular or bone malformation, and also for thoracic and digestive pathologies (Fig. 1). Thanks to direct volume rendering, it will be easier to detect the pathology, compute distance and see the related link with surrounding organs such as muscles.

Volume rendering is thus a very useful tool since it is accessible without any preprocessing, but has also some limitations. Mainly, it can neither provide the volume of structures nor their dimensions since these structures are not delineated. For the same reason, it is not possible to provide a volume after resection or to cut only a part of the structures without cutting neighboring structures. To overcome this limit, each anatomical and pathological structure in the medical image has to be delineated. Such delineation is a long and difficult manual task for radiologists with a standard software. Therefore, several research teams have developed automated

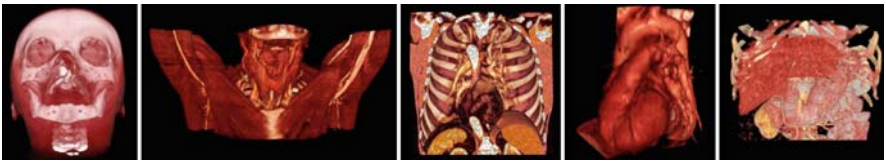


Fig. 1 VR-Render direct volume rendering on CT-scan of head, thyroid, thorax, heart and abdominal area

Fig. 2 3D modeling of organs from head to pelvis using VR-Anat

3D patient modeling software that can interactively provide the main visible organs from a CT-scan or MRI [3–7]. We have also developed several 3D patient modeling software applications, which automatically or interactively provide the main visible organs from a CT-scan or MRI for the digestive area [8,9]. Our latest software, VR-Anat ©*IRCAD 2009*, allows to perform 3D modeling from the head to the pelvis area, thanks to an improved management of medical knowledge combined with an interactive process [10–12] (Fig. 2).

Our method is based on a hierarchical segmentation of visible organs, from the simplest to the most complex one. The first step detects and delineates respectively skin, lungs, aorta, kidneys, spleen and bones. Each organ segmentation allows for reduction of the image to a smaller one containing other, not yet segmented organs. Moreover, segmented organs can be used to provide useful information such as location of not yet segmented organs.

We first start with a thresholding step fixed between −1024 and −300 HU. It allows to extract several components corresponding to the air around skin, air in lungs and air in the abdomen. A simple morphological opening followed by an analysis of connected components then allows to easily extract skin and lung frontiers. Indeed, external air is the biggest component and lungs are the two major components in the upper slices. Bone extraction is a more complex task, contrasted vessels, some internal calcified structures and part of kidneys appearing as bright as the lowest bone densities. Kidneys and aorta will then be segmented before.

The aorta is extracted from its density and its shape. Indeed, the aorta is a big vessel coming from the heart and following vertebrae, so we first detect the vertebrae location. We extract the major 3D components obtained after a thresholding over 250 HU followed by a 3D morphological dilation with a radius of 1 cm. Thus, we obtain a set of connected bones and vessels with no hole between vertebrae. We then perform a 3D morphological erosion with a radius of 1.6 cm allowing to remove all 3D components with a radius under 3.2 cm (aorta, ribs). The main 3D component is necessarily a kind of tube corresponding to the middle of the vertebrae. After a dilation with a radius of 8 mm (2 mm over the initial size obtained with the thresholding), we finally obtain an approximation of vertebrae localization. On each slice, size (dx_z, dy_z) and localization ($x_{\min z}$, $y_{\min z}$) of vertebrae area are then computed. Reducing the image to the area defined from ($x_{\min z}$, $y_{\min z} - dy_z$) to ($x_{\min z} + 3/2dx_z$, $y_{\min z} + dy_z$) for each slice z, we obtain a reduced area that contains the aorta.

We then keep all voxels with a density contained between 120 and 350 HU, which ensures that the aorta will always appear in the thresholded result. Of course, such a

fixed value does not take variations between patients into account, and in most cases, the thresholded result provides surrounding veins and part of the bones, sometimes connected to the aorta. To disconnect the aorta from these surrounding structures, we perform a 2D morphological erosion with a 4 mm radius. We then extract 2D connected components in each slice and analyze the shape of each component with a method used in Gao et al. [13] to analyze the 2D tumor shape. Indeed, the aorta being perpendicular to the axial slices, it appears like a circle in slices. All circle structures are then kept in the image, corresponding to possible vessels, other structures being removed. On several slices, junction between artery and aorta can result in a 2D shape far from a circle. On such slices, aorta is then removed and not recognized. But, even if some parts of the aorta could thus be removed, the biggest 3D component will be the aorta. From this biggest component, we perform a simple region growing on the thresholded and eroded image allowing to extract the aorta. Then, a simple morphological dilation with a 4 mm radius provides the whole aorta. Smallest arteries (such as kidneys, mesenteric and femoral arteries) can then be segmented from the aorta through a simple region growing using the min and max threshold extracted from segmented aorta.

The kidney delineation method is based on the gray-level histogram analysis of the image limited to regions including kidneys. Indeed, the right inferior quarter of the image essentially contains a part of the liver and the right kidney, whereas the left inferior quarter of the image only contains the right kidney and the spleen, and the half upper image essentially contains the liver. A comparative analysis of grey-level histograms then allows to find the intensity range of kidneys, localized identically on both histograms, and of the spleen located in the lower left quarter.

In each sub-image, we keep the main connected component composed of voxels with a density under 250 HU and over the valley minimum between both maximum picks. After a morphological opening and closing, we obtain two kidneys. The spleen is then extracted from a first thresholding performed on the half left part of the image. Threshold values are computed around the spleen grey-level picks in the histogram, when the slope is inferior to 10 (beginning of a valley). Such a threshold provides spleen and liver in the same half left image. We then extract both main connected components after a morphological opening with a 4 mm radius. The spleen is then always the component located on the left of the liver and in the lowest position.

We finally extract bones using a distance map of the skin, bones being characterized close to the skin. Each 200 HU thresholded connected component is then labeled with the minimum distance to the skin. The component with the smallest label *Dmin* is necessarily a bone. In order to extract all bones, we then keep all connected components with a label lower than the *Dmin* +1 cm. Such a method provides all bones (ribs and vertebrae).

Skin, aorta, bones, lungs, kidneys and spleen are thus automatically delineated using thresholding, gradient, shape and morphology-based methods. All defined constraints are based on simple localization and shape rules used by radiologists and surgeons when they analyze images. From this segmentation, delineated organs are then removed from the initial image which allows to reduce the image to a smaller

one containing other not yet segmented organs. Moreover, segmented organs can be used to provide useful information such as location for the segmentation of not yet segmented organs. The second step then performs the delineation of veins, liver and its internal tumors with our method [8] or by using more recent techniques [14, 15].

Several companies propose today such 3D modeling as an online distant service by delivering a DICOM image essentially for dental pathologies, but also more recently for the digestive area (MeVis Distant Services AG, PolyDimensions GmbH). In the same way, since 2003 we have set up several free cooperations with distant university hospitals (Geneva, Montreal, Lausanne and Strasbourg) in order to offer a similar experimental 3D modeling service called MEDIC@. More than 800 clinical cases have thus been modeled in 3D for the thoraco-abdominal area, 60% being reconstructed for liver pathologies. As the MeVis Distant Services, this distant 3D modeling service showed its efficiency to provide an accurate 3D model of patient anatomy and pathology in a short delay (between 1 and 3 days). Such web-based services should represent in the future the first step of any computer-assisted endoscopic surgery.

Existing 3D delineation software platforms are usually linked to surgical planning tools developed by the same teams [16–19]. In order to exploit our 3D modeling, we have developed a surgical planning system (3D VSP). Beside the 3D visualization of delineated and modeled structures, this system allows to put each structure into transparency, to interact on them, to navigate anywhere and therefore to simulate any kind of endoscopy: laparoscopy, fibroscopy, gastroscopy, colonoscopy and transvaginal NOTES (Fig. 3). All these endoscopic simulations represent a great improvement in NOTES planning by providing a preoperative patient specific information that will help surgeons during the real procedure.

It further allows to perform virtual resections defined by interactively positioned cutting planes, to provide the volume of all visualized structures and to plan minimally invasive tool positioning (Fig. 4). Because of its compatibility with current standards, this system can be used on a laptop fitted with a 3D graphic card and can therefore be used during the intervention so as to improve the control of the gesture.

Pediatric pulmonary sequestration represents a good illustration of benefits obtained with such a tool. Indeed, this pathology, located on the aorta of a newborn baby, can be cured only through a difficult minimally invasive approach. In such clinical cases, after 3D modeling of a patient with VR-Anat, 3D VSP will allow to perfectly plan the surgical tool positioning by providing a virtual laparoscopic view similar to the real surgical view (Fig. 5). It also allows virtual artery ligature

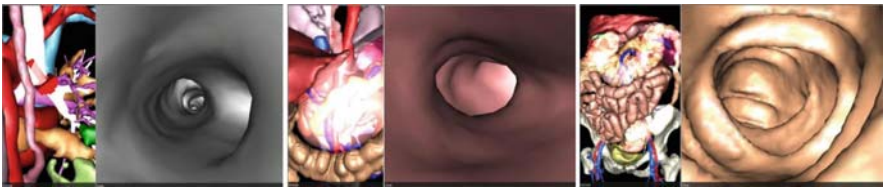


Fig. 3 Patient-specific virtual fibroscopy (*left*), gastroscopy (*centre*) and colonoscopy (*right*)

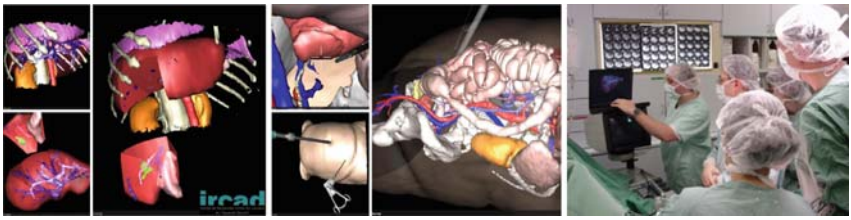


Fig. 4 Virtual resection of a 3D reconstructed liver from medical images (*left*), virtual laparoscopic tool positioning in uterus area (*centre*) and intra-operative clinical use of the planning software on a laptop (*right*)

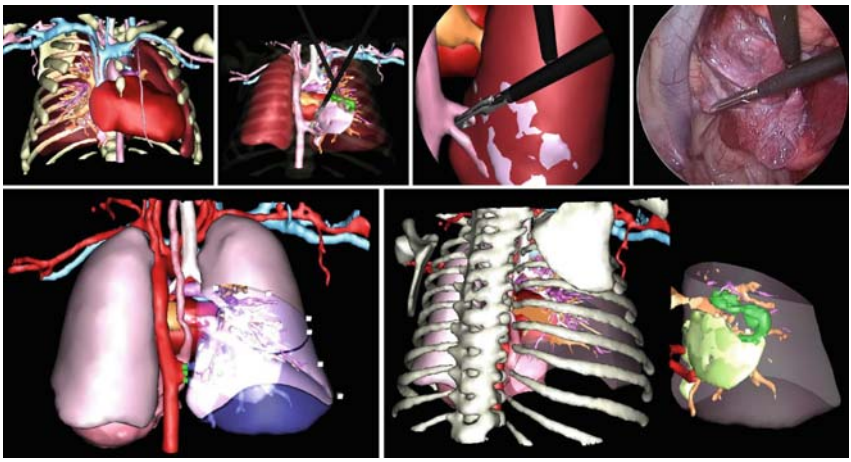


Fig. 5 3D modeling of a 3 month-old baby and the associated preoperative surgical planning provides a perfect tool placement (*first line*) and resection definition (*second line*) that will be reproduced intraoperatively for safer surgery

and lung resection providing postoperative lung volume and a better visualization of pulmonary arteries and veins that will be cut during this resection. This virtual resection is performed interactively, what provides an improved surgical procedure definition and finally, safer surgery.

Virtual Reality for Preoperative Surgical Simulation

Surgical planning allows to define the best operative strategy, but it does not provide the possibility to evaluate a surgeon's ability to perform this strategy. This is the objective of realistic surgical simulators. To be efficient, these simulators have to reproduce sensations linked to operative gestures: realistic visual perception and realistic haptic rendering through force feedback. Force feedback is an important issue

for any realistic surgical simulator. Today, there is a large range of haptic systems enabling reproduction of free-hand movements of instruments such as a scalpel or needle (essentially Phantoms from Sensable), or constrained instrument movements such as laparoscopic tools (Laparoscopic Impulse Engine from Immersion, Instrument Haptic Port from Xitact, Force Feedback from Karl Storz). There is today a large number of commercial surgical simulators for abdominal surgery, urology, gynecologic or arthroscopic procedures (Surgical Science, Simbionix^{TM}, SimSurgery, Haptica or Karl Storz). Various scenarios are then available as separate modules. The increasing realism of visual rendering, due to the use of textures obtained from real images, and the progress in force feedback mechanisms enabled those products to acquire some maturity. Though they are attractive, these simulators are limited to the simulation of restricted and determined virtual models within a set database. Moreover, no system is dedicated to pediatric surgery.

The main idea of a patient-specific simulator is to offer surgeons the opportunity to carry out preclinical training on a virtual copy of the patient. Only one company (Simbionix^{TM} in partnership with PolyDimensions GmbH) proposes today such a preoperative simulation but limited to arthroscopy. Few research works propose also a preoperative patient-specific simulation, essentially on endoscopic procedures [20, 21] or orthopedic surgery [22–24], and sometimes on other more complex procedures [25]. To meet those limits, we have developed a first patient-specific simulator, the Haptic Operative Realistic Ultrasonographic Simulator called HORUS [26]. HORUS is dedicated to ultrasound-guided percutaneous interventions for training or for preoperative simulation from CT or MRI images of patients. To simulate a realistic ultrasonographic image, we have developed an automatic process that converts a 3D volumetric medical image (CT-scan or MRI) into a realistic US image [27] and simulated the thermal ablation effect (Fig. 6). The system has successfully been tested with a 1 mm CT-scan of a fetus at 36 weeks gestation, a 2 mm MRI image of a fetus at 29 weeks gestation, four patients for hepatic biopsy and two patients for hepatic thermal ablation.

From HORUS, we developed in partnership with the Karl Storz Company a laparoscopic simulator called ULIS (Unlimited Laparoscopic Immersive Simulator). The objective of its first version is to be the first patient-specific laparoscopic basic skill training simulator using the Karl Storz Force Feedback System allowing to introduce real surgical instruments. This simulator uses 3D modeled patients extracted from CT-scan reconstructions that can be deformed locally with force feedback

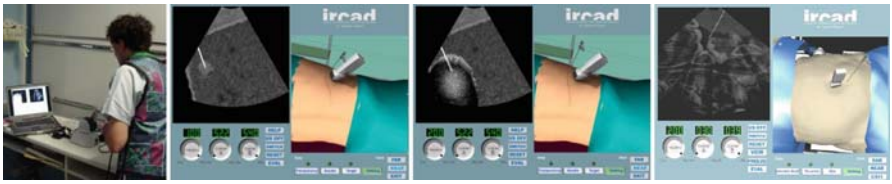


Fig. 6 HORUS allows a preoperative patient-specific simulation of thermal ablation, the burning effect being also simulated in the virtual ultrasonographic image (*right*)

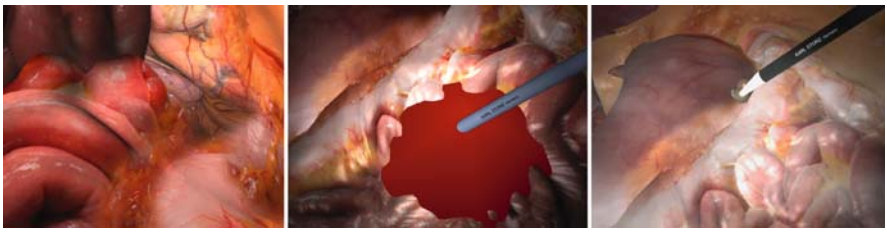


Fig. 7 Laparoscope manipulation (*left*), virtual blood injury in the bowel area (*centre*) and coagulation including smoke rendering (*right*) in ULIS

rendering when a virtual instrument is in contact with tissues. A photo-realistic texturing of tissues is performed manually, which required more than a week of work. ULIS includes camera and surgical instrument manipulation, and simulation of suction and coagulation procedures (Fig. 7). The next step of this work will consist in developing a preoperative version, which means an automatic texturing of 3D modeled organs and the integration of the complete simulator engine SOFA [28] (Simulation Open Framework Architecture), allowing multiple interactions and including resection, developed in our first patient-specific simulator for laparoscopic liver surgery. This work will be performed in the PASSPORT project, funded by the ICT eHealth program of the European Community in the seventh Framework Programme. It will then be adapted to NOTES by integrating the flexible tools and the endoscopic simulation.

Virtual Reality for Intra-operative Assistance: Augmented Reality

Preoperative surgical planning and simulation will mainly improve the efficiency of MIS procedures thanks to a better preoperative learning of the patient's anatomy and a real preoperative training. But preoperative use of virtual reality is not sufficient to insure safety during the surgical procedure. Such an improvement can be provided by an intra-operative use of virtual reality through the Augmented Reality (AR) concept. Indeed, AR consists in superimposing the preoperative 3D patient modeling onto the real intra-operative patient view. AR provides a transparency view of patients and can also guide surgeons, thanks to the virtual improvement of their real surgical tools that are tracked in real-time during the procedure. We have developed two kinds of AR software applications: Interactive Augmented Reality and Fully Automatic Augmented Reality.

Interactive Augmented Reality [29] is composed of two steps. The first step consists in registering the virtual patient view with the real patient view. The second step consists in moving virtual tools in real-time, putting them in the same position

Fig. 8 Interactive Augmented Reality on liver (*left*), pancreas (*centre*) and parathyroid (*right*) video-assisted surgery

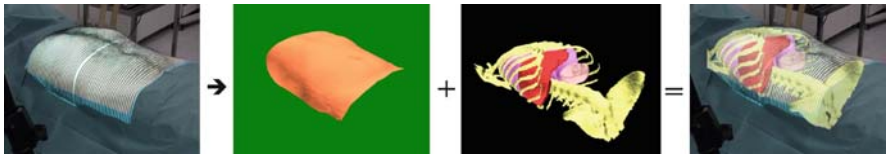


Fig. 9 Use of structured light combined with breathing movement simulation for the real-time augmented reality of organs in movement without any skin markers

as the real ones. These two steps being performed manually using an interactive software, which allows to correct all visible mistakes or bad registrations (Fig. 8).

Such an augmented reality is necessarily user-dependant due to manual positioning. Therefore, it is impossible to ensure accuracy for this first interesting image-guided surgical tool that can only be used for an easier understanding of patient anatomy using pre-operative data. This limitation can be overcome by automatically superimposing preoperative data onto the real patient during the intervention and by automatically tracking surgical tools in real-time.

Automatic registration is the first problem to solve. Indeed, preoperatively, during the CT acquisition, the patient holds his/her breath which is not the case during the surgical procedure. Organ movements due to breathing create the difference between the patient-specific model and the real organ during surgery. To solve this problem, we have developed a predictive deformation of organs and tissues during a breathing cycle from a single CT image and a video sequence of the patient. Results are very encouraging and show the feasibility of such a predictive simulation control by the real breathing of the patient [30]. They are due to the use of volume deformation fields that are defined from skin movements tracked thanks to the use of structured light and to the preliminary segmentation of thorax-abdomen regions and organs. The use of two cameras allows then an efficient surface reconstruction of the skin, but also provides a real vision of the patient on which we can superimpose the virtual model of the patient (Fig. 9). Registration is then performed using our 3D/2D registration method [31].

Our objective being to use such a system during laparoscopic and NOTES surgery, our next work will consist in adapting this system by taking the deformation due to the insufflation of gas inside the abdomen into account.

Automatic surgical instrument tracking is a second problem to be solved. Today, there is a large set of tracking systems, essentially based on optics and electromagnetic fields. In the laparoscopic application, we choose not to use a second system by using the same two cameras used for patient registration to track instruments [32]. To do so, we have adapted the ARToolKit library [33] so that it works with USB or Firewire cameras. Thus, tracking of the instrument is eased by adding a printed planar square, which is automatically recognized by the library, and permanent visualization of tool positioning is also possible (Fig. 10).

Our validation, performed onto the same synthetic liver model, shows a 0.6 mm registration accuracy [33]. We have thus obtained a fully efficient and accurate image-guided surgical tool for abdominal procedures that we have tested in vivo in a preclinical evaluation step. This evaluation was performed on five rats. The medical imaging was realized on a Siemens-CTI microscanner MICRO CATII used in IRCAD to perform in vivo follow-up of tumor evolution in fundamental research in the field of digestive cancer. To visualize anatomical structures, we have injected intravenously 3 ml of LC Fenestra contrast medium (©ART) 30 min before realizing the pre-operative CT-scan. As on human experimentation performed in interactive augmented reality, fiducials positioned on the skin before performing the CT-scan are used to do the automatic registration (Fig. 11).

Fig. 10 Fully automated augmented reality integrating a real-time tracking of tools

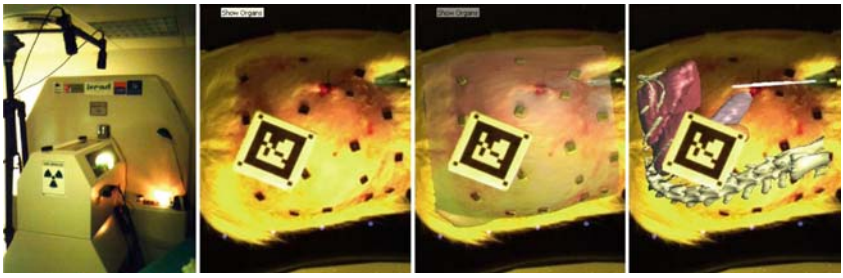


Fig. 11 Preclinical validation on rats is performed from an in vivo micro-CT scan (*left*). The Automatic Augmented Reality provides a virtual transparency of anatomical structures thanks to fiducial registration

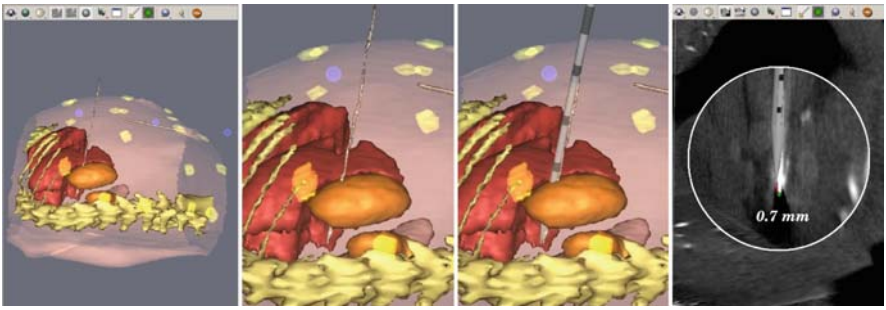


Fig. 12 Preclinical validation on rats shows a 0.7 mm accuracy between the augmented reality needle and its real position extracted from the control CT-scan (*right*)

To evaluate the efficiency of the global system, we had then inserted an automatically tracked needle in the right kidney of each rat. This needle is then visualized on a control CT-scan. The comparison between the real positioning of the needle and the augmented reality visualization of this needle is then possible and shows a precision of 0.7 mm (Fig. 12). The next work will consist in testing it clinically.

For NOTES surgery, optical tracking is not sufficient. Thus, we have developed and patented a new automatic tracking system called METRIS 3D, providing in real-time the 3D shape and location of the flexible endoscope. This system is based on the Aurora tracking system developed by NDI. It is composed of a 1.20 m long tube incorporating several miniature electromagnetic coils distributed on its length and which can be located by an Aurora tracking system. It can thus be introduced into the operating channel of a flexible endoscope. Locating these various coils like a GPS, the software developed in link with that device allows both to reconstruct the tube shape and location in three dimensions. After a calibration phase, and thanks to the tracking of the endoscopic camera, it is then possible to provide a real-time automatic augmented reality endoscopic view that can be used for endoscopy or NOTES.

Conclusions

We presented a set of tools for diagnosis support and surgical intervention planning. They also allow to use preoperative information during the intervention, in particular by superimposing the virtual image of internal organs and pathologies onto the abdomen of the patient. These systems, at an experimental stage, are progressively being tested clinically for endoscopic and NOTES surgery, with the objective of eventually being used in clinical routine. Our first results show that the systems works very efficiently if organs do not move or if there is no other movement than breathing induced movements. But they remain ill-adapted for moving organs in

a surgical procedure where surgeons modify interactively the shape, position and topology of organs. To solve this problem, we want to combine in the future a realistic, even predictive, simulation with the augmented reality system. But to be fully efficient, this system will require to gather intra-operative information in real-time so as to control the simulator. In this respect, we are working in close cooperation with the EAVR team of the LSIIT, led by Michel De Mathelin within the IRMC consortium that pools five teams of the University of Strasbourg. These works have therefore been allowed to automatically filter respiratory movements and heartbeat movements via visual servoing of a surgical tool-holding robot [34]. They have also permitted to obtain automated suturing [35] and automated needle positioning for percutaneous thermal ablation [36].

These works represent the first essential phase for surgical gesture automation, which will allow to reduce surgical mistakes. Indeed, intervention simulation will allow to do without all superfluous or imperfect gestures, using it as a programming of the final gesture. This gesture will then be transmitted to a surgical robot which, thanks to augmented reality and visual servoing, will be able to precisely reproduce optimized gestures of the surgeon. Tomorrow's surgery is on its way.

Acknowledgements These works are a part of the European eHealth project called PASSPORT, which is funded by the European Community within the ICT theme of the seventh Framework Programme.

References

1. OsiriX Imaging Software: www.osirix-viewer.com
2. VR-Render: www.ircad.fr/softwares/vr-render
3. Seong W, Kim E-J, Park J-W (2004) Automatic Segmentation Technique Without User Modification for 3D Visualization in Medical Image CIS 2004. Lect Notes Comput Sci 3314:595–600
4. Schenk A, Prause G, Peitgen H-O (2000) Efficient semiautomatic segmentation of 3D objects in medical Images MICCAI 2000. Lect Notes Comput Sci 1935:186–195
5. Young Y-N, Levy D (2006) Registration-based morphing of active contours for segmentation of CT scans. Mathematical Biosciences and Engineering 2(1):79–96
6. Camara O, Colliot O, Bloch I (2004) Computational modeling of thoracic and abdominal anatomy using spatial relationships for image segmentation. Real Time Imaging 10(4): 263–273
7. Kitasaka T, Ogawa H, Yokoyama K, Mori K, Mekada Y, Hasegawa J-I, Suenaga Y, Toriwaki J (2005) Automated extraction of abdominal organs from uncontrasted 3D abdominal X-ray CT images based on anatomical knowledge. J Computer-Aided Diagn Med Images 9(1):1–14
8. Soler L, Delingette H, Malandain G, Montagnat J, Ayache N, Koehl C, Dourthe O, Malassagne B, Smith M, Mutter D, Marescaux J (2001) Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery. Comput Aided Surg 6(3):131–142
9. Simone M, Mutter D, Rubino F, Dutson E, Roy C, Soler L, Marescaux J (2004) Three-dimensional virtual cholangioscopy: a reliable tool for the diagnosis of common bile duct stones. Ann Surg 240(1):82–88
10. Fasquel JB, Agnus V, Lamy J (2006) An efficient and generic extension to ITK to process arbitrary shaped regions of interest. Comput Methods Programs Biomed 81(1):1–7

11. Fasquel JB, Agnus V, Moreau J, Soler L, Marescaux J (2006) A medical image segmentation system based on the optimal management of the regions of interest using topological medical knowledge. Comput Methods Programs Biomed 82(3):216–230
12. Fasquel JB, Brocker G, Moreau J, Agnus V, Papier N, Koehl C, Soler L, Marescaux J (2006) A modular and evolutive component oriented software architecture for patient modelling. Comput Methods Programs Biomed 83(3):222–233
13. Gao L, Heath DG, Kuszyk BS, Fishman EK (1996) Automatic liver segmentation techniques for three-dimensional visualization of CT data. Radiology 2(201):359–364
14. Lim S-J, Jeong Y-Y, Ho Y-S (2006) Automatic liver segmentation for volume measurement in CT Images. J Vis Commun Image Represent 17(4):860–875
15. Young Y-N, Levy D (2006) Registration-based morphing of active contours for segmentation of CT scans. Math Biosci Eng 2(1):79–96
16. Meinzer HP, Schemmer P, Schöbinger M, Nolden M, Heimann T, Yalcin B, Richter GM, Kraus T, Büchler MW, Thorn M (2004) Computer-based surgery planning for living liver donation. 20th ISPRS Congress, Istanbul 2004, International Archives of Photogrammetry and Remote Sensing. XXXV(B):291–295
17. Numminen K, Sipilä O, Mäkisalo H (2005) Preoperative hepatic 3D models: virtual liver resection using three-dimensional imaging technique. Eur J Radiol 56:179–184
18. Radtke A, Nadalin S, Sotiropoulos GC, Molmenti EP, Schroeder T, Valentin-Gamazo C, Lang H, Bockhorn M, Peitgen HO, Broelsch CE, Malago M (2007) Computer-assisted operative planning in adult living donor liver transplantation: a new way to resolve the dilemma of the middle hepatic vein. World J Surg 31:175–185
19. Koehl C, Soler L, Marescaux J (2002) A PACS based interface for 3D anatomical structures visualization and surgical planning. SPIE proceeding 4681:17–24
20. Chung A, Deligianni F, Shah P, Wells A, Yang G-Z (2004) Enhancement of visual realism with BRDF for patient specific bronchoscopy simulation. Medical Image Computing and Computer-Assisted Intervention, Lect Notes Comput Sci 3217:486–493
21. Suzuki S, Eto K, Hattori A, Yanaga K, Suzuki N (2007) Surgery simulation using patient-specific models for laparoscopic colectomy. Stud Health Technol Inform 125:464–466
22. Agus M, Giachetti A, Gobbetti E, Zanetti G, Zorcolo A (2002) A multiprocessor decoupled system for the simulation of temporal bone surgery. Comput Vis Sci 5(1):35–43
23. Cohen ZA, Henry JH, McCarthy DM, Mow VC, Ateshian GA (2003) Computer simulations of patellofemoral joint surgery. Patient-specific models for tuberosity transfer. Am J Sports Med 31(1):87–98
24. Luboz V, Chabanas M, Swider P, Payan Y (2005) Orbital and maxillofacial computer aided surgery: patient-specific finite element models to predict surgical outcomes. Comput Methods Biomech Biomed Engin 8(4):259–265
25. Sierra R, Dimaio SP, Wada J, Hata N, Székely G, Kikinis R, Jolesz F (2007) Patient specific simulation and navigation of ventriculoscopic interventions. Stud Health Technol Inform 125:433–435
26. Soler L, Forest C, Nicolau S, Vayssiere C, Wattiez A, Marescaux J (2007) Computer-assisted operative procedure: from preoperative planning to simulation. Eur Clin Obstet Gynaecol 2:201–208
27. Hostettler A, Forest C, Forgione A, Soler L, Marescaux J (2005) Real-time ultrasonography simulator based on 3D CT-scan images. Stud Health Technol Inform 111:191–193
28. Allard J, Cotin S, Faure F, Bensoussan P-J, Poyer F, Duriez C, Delingette H, Grisoni L (2007) SOFA – an open source framework for medical simulation. Stud Health Technol Inform 125:13–18
29. Marescaux J, Rubino F, Arena M, Soler L (2004) Augmented reality assisted laparoscopic adrenalectomy. JAMA 292(18): 2214–2215
30. Hostettler A, Nicolau S, Soler L, Remond Y (2006) Real time simulation of organ motions induced by breathing: first evaluation on patient data ISBMS. Lect Notes Comput Sci 4072: 9–18

31. Nicolau S, Pennec X, Soler L et al (2003) Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by augmented reality IS4TM. Lect Notes Comput Sci 2673: 270–283
32. Nicolau S, Goffin L, Soler L (2005) A low cost and accurate guidance system for laparoscopic surgery: validation on an abdominal phantom VRST. Proceeding of ACM Symposium on Virtual Reality Software and Technology, pp 124–133
33. Nicolau S, Garcia A, Pennec X, Soler L, Ayache N (2005) An augmented reality system to guide radio-frequency tumor ablation. J Comput Animation Virt World 16(1):1–10
34. Krupa A, Gangloff J, Doignon C, de Mathelin M, Morel G, Leroy J, Soler L, Marescaux J (2003) Autonomous 3D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE transactions on robotics and automation 19(5):842–853
35. Nageotte F, de Mathelin M, Doignon C, et al (2004) Computer-aided suturing in laparoscopic surgery. Proceeding of Computer Assisted Radiology and Surgery, Chicago, USA, June 2004
36. Maurin B, Piccin O, Bayle B, et al (2004) A new robotic system for CT-guided percutaneous procedures with haptic feedback. Proceedings of Computer Assisted Radiology and Surgery, Chicago, USA, June 2004

Design of a Robotized Flexible Endoscope for Natural Orifice Transluminal Endoscopic Surgery

Bérengère Bardou, Florent Nageotte, Philippe Zanne, and Michel de Mathelin

Introduction

Natural orifice transluminal endoscopic surgery (NOTES) is a new surgical technique, which consists in reaching the peritoneal cavity through a natural orifice (mouth, anus, or vagina) in order to treat a specific area (Fig. 1). It allows to perform various procedures such as cholecystectomy or tubal ligation. The main advantage of this kind of surgery is the absence of visible scars since it does not need any external incision. Thanks to this procedure, postoperative pain, recovery time, and psychological impact are reduced. The first transluminal procedure has been performed by Pr. Marescaux in Strasbourg in 2007 [1].

The equipments currently used for NOTES are flexible endoscopes from gastroenterology. These endoscopes are composed of a long flexible shaft, more than 1 m long, with an articulated bending tip. The bending tip is about 10 cm and can move along two perpendicular directions. The tip is linked to navigation wheels thanks to cables running along the shaft of the endoscope. The surgeons control the motion of the tip by turning the navigation wheels. Visual feedback is provided by a miniature camera embedded at the tip.

Transluminal procedures are challenging for surgeons. Indeed the position and the shape of the flexible shaft of the endoscope cannot be controlled. The shape is defined by the anatomical structures in contact with the endoscope. The surgeons do not know this shape and every movement of the body of the endoscope leads to unpredictable effects in the images. Moreover, the lack of degrees of freedom (DOFs) limits the use of classical endoscopes to a few basic tasks, while complex tasks, such as stitching or knot tying, are very difficult or even impossible to perform. Hence, surgeons need additional degrees of freedom and better manipulability. That is why the endoscopic system has been redesigned to add articulated arms, which provide at least two additional DOFs each. With this new kind of system, the workspace is increased and surgeons can perform more complex surgical tasks in a more intuitive way.

B. Bardou (✉), F. Nageotte, P. Zanne, and M. de Mathelin
LSIIT laboratory, University of Strasbourg, CNRS, France
e-mail: bardou.berengere@unistra.fr

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_9,

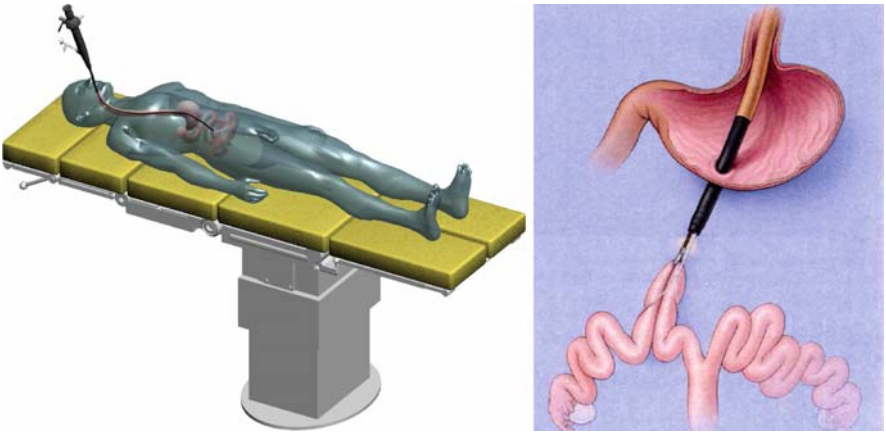


Fig. 1 Principle of transluminal surgery

Problematics

Currently available flexible endoscopes are inadequate for performing complex transluminal surgical procedures. Issues with conventional scopes include:

- The insufficient triangulation angle between the surgical instruments
- The lack of a multitasking platform which allows the operator to perform several tasks independently, such as holding an instrument and grasping an organ
- The limited number of access channels for driving surgical instruments to the operation area
- The limited number of available DOFs
- The inability to properly control surgical instruments

A long term solution would be to redesign the flexible endoscopic system. However, flexible endoscopes are widely used systems which allow to overcome several difficult problems such as cleaning, asepsis, and good visual feedback. Therefore they are a good basis for developing more complex systems. Several prototypes have been proposed and are currently tested [2]. The design which has been widely chosen consists of a video endoscope and two arms. The arms are attached to or included in the main endoscope. In all systems, the distal DOFs can be controlled from the proximal end of the system using cables. Other difficulties, consequences of the previously mentioned problems, have arisen during surgical procedures. Indeed, once the target has been reached, the surgeon must combine several actions to achieve adequate motions. Given the high number of DOFs, several surgeons must cooperate for carrying out a task (see Fig. 2). They share a small workspace and must show a good coordination to perform the operation.

Fig. 2 Experiments carried on at the IRCAD in Strasbourg using a manual transluminal system

Related Work and Systems

In the following, we briefly review some of existing systems and their specificities. The nonexhaustive list of systems we compare is made-up of systems developed in university laboratories (the endoscopic robotic system from Jikei University, Tokyo-Japan [3], the Nanyang University endoscopic system, Singapore [4], and the HVSPS from the Technical University in Munich [5]) and industrial prototypes (the ViaCath from HansenMedical [6], the Cobra system from USGI Medical, the Anubiscope from Karl Storz endoscopy which has been very recently presented at the SAGES congress, the system proposed by Olympus [7] and the multiendoscopes system from Pentax [8]). They can be classified according to several criteria.

Use of an Overtube

Several systems use an overtube with three channels [Cobra, Tokyo University, ViaCath, Pentax, Nanyang University, Munich University]. One channel is used for passing a flexible endoscope which brings light and visual feedback, and the other channels are used for instruments. This overtube is generally passive and flexible. However the overtube of the ViaCath system can be controlled along two directions at the tip. The overtube of the Cobra system is flexible but it can be locked in shape.

For the other systems [Storz, Olympus] the light and visual feedback is directly brought by the main system. There are only two channels in the system for passing instruments and the tip of the main endoscope is controlled.

Articulated or Passive Instruments

A key difference between the systems concerns the use of specific articulated instruments or conventional passive instruments. In most systems [Tokyo University, ViaCath, Pentax, Nanyang University, Munich University], a part of the DOFs is directly provided by the surgical instruments through discrete or snake-like joints. However, in the Olympus and Storz systems, instruments are passive. A part of the DOFs is brought by hollow arms. The surgical tools are inserted inside the arms, which are in turn inserted inside the channels of the main system [Storz] or directly fitted at the tip of the main endoscope [Olympus].

Discrete or Continuous DOFs

Most of the systems use snake-like continuous DOFs as in classical endoscopes. However, discrete DOFs are used for the instruments in the Nanyang University and in the ViaCath system. These are also actuated using cables as in continuous snake-like systems.

The Triangulation Problem

The triangulation of the instruments is a key feature for good operating conditions. Triangulation is generally brought by deflecting the instruments from the main endoscope direction [Olympus, Munich, Tokyo, Pentax]. In the ViaCath system, the triangulation is provided by the multiple DOFs of the instruments. The Storz prototype proposes an original solution based on the use of articulated flaps at the tip of the endoscope. Closed flaps allow the insertion of the endoscope into the human body. When opened, the flaps deflect the tip of the channel and hence the surgical instruments, thus allowing the triangulation of video and arms (see Fig. 3). Pentax also offers a solution to the triangulation problem by deflecting the small endoscopes when they come out of the channels of the overtube.

Manipulation and Robotization of the Systems

Specific master consoles have been proposed to manipulate these complex devices. They allow for a single surgeon to control the whole slave system and to perform the surgical task. In order to make the manipulation of the system easy and intuitive, Olympus offers a control interface which is similar to manual 4 DOFs laparoscopic surgical instruments and allows to mechanically drive the arms.

Fig. 3 Karl Storz system and its control handles

But among the presented systems, only the ViaCath and the Singapore systems have been motorized. An exoskeleton interface has been proposed to easily control all the DOFs of the Singapore system.

Presentation of our Endoscopic Prototype

We have developed a new endoscopic system. The main objective is to propose a teleoperated system that allows a single surgeon to perform complex NOTES procedures. Our prototype solves the triangulation problem thanks to a special device and active flexible arms. The robotization of the whole system improves the instruments manipulability and is a key element to stabilize the surgeon movements.

The Mechanical System

The slave system has been developed in collaboration with Karl Storz[1] based on the manual system briefly described in the previous section. Our prototype consists of two flexible hollow arms which are attached to a conventional flexible endoscope with 2 DOFs (1 and 2 on Fig. 4). The endoscope provides the optical system for visual feedback and two working channels for conventional instruments. The hollow arms are fixed on the circumference at the end-part of the bending tip of the endoscope using a specific cap. Each arm provides 2 DOFs similarly to the main endoscope: left/right and up/down motions (5 and 6). Surgical instruments can be introduced into the arms and guided to the operating area. These instruments can translate and rotate inside the arms (7 and 8). The cap attached to the endoscope also allows to deflect the arms from the main direction of the endoscope. This provides

[1] Karl Storz Endoskope GmbH, Germany.

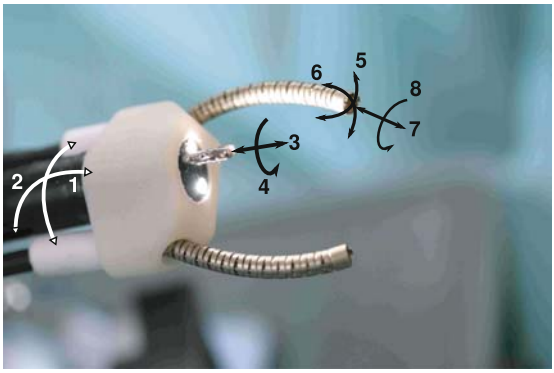


Fig. 4 The tip of the prototype with the special end cap and the DOFs

triangulation between the arms and the endoscopic view, it enhances the cooperation area of the two arms, and makes the use of the system more intuitive. A view of the tip of the prototype is presented in Fig. 4.

Motorization

The orientation of the main endoscope as well as the orientation of both arms is driven by two cables which run through the flexible shaft up to the proximal end where they are rolled up around pulleys.

These pulleys are controlled by rotary motors mounted on the endoscope and the arms handles (see Fig. 5). So as to choose the solution for the motorization, the torques required to bend the tip of the main endoscope and the arms have been estimated. The results of the torque assessment are given in Table 1. The motor which drives the main endoscope orientation must also be able to bend the passive flexible body of the arms.

The smallest available motors allowing to reach these torques have thus been chosen off-the-shelf. These are Harmonic Drive FHA-8C hollow shafts motors which allow up to 1.5 Nm torque for the main endoscope (shown on Fig. 6) and Harmonic Drive RSF for the arms.

The instruments are translated into channels and into arms thanks to linear motors. Moreover, a rotary motor can be mounted onto the translation unit so as to rotate the surgical instruments around their own axis (see Fig. 7).

Finally, the whole system can have up to 12 actuated DOFs which can be combined in several manners and can hence provide solutions to many medical procedures.

All motors are velocity controlled at low-level with external position loops running at 2,000 Hz on a unique controller. A low bandwidth loop running at the frame rate of the endoscopic camera can be added for providing image-based control of the system.

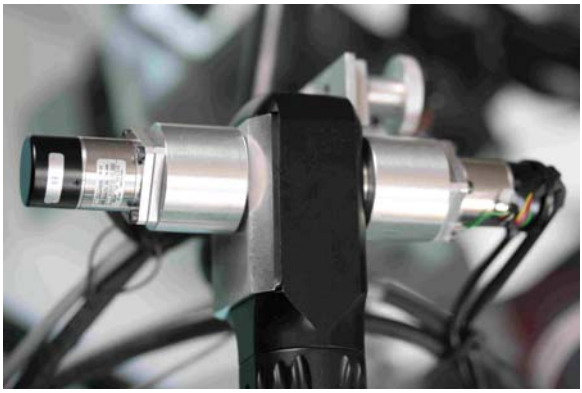


Fig. 5 Motorization of the arms

Table 1 Choice of the motorization solution

	Max req. torque (Nm)	Motor	Max. continuous mot. torque (Nm)
Main endoscope	1.3	HD FHA	1.5
Hollow arms	0.20	HD RSF	0.45

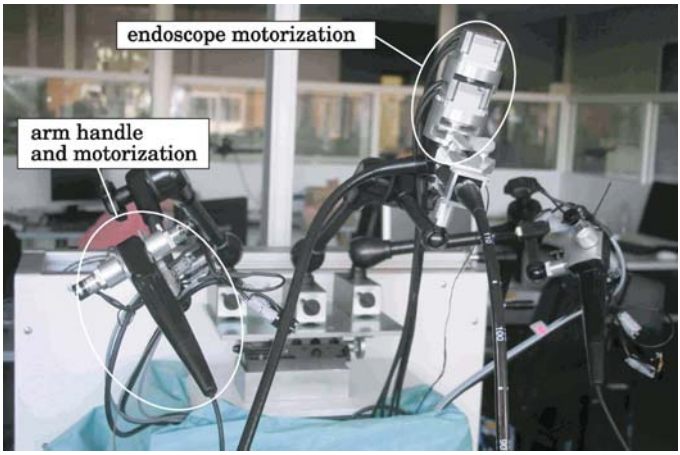


Fig. 6 Motorization

Master Control of the System

The surgeon console looks like a desk workstation. It carries two master interfaces and two flat panel monitors for displaying the images of the endoscope and other visual information. The master interfaces are Omega interfaces by ForceDimension

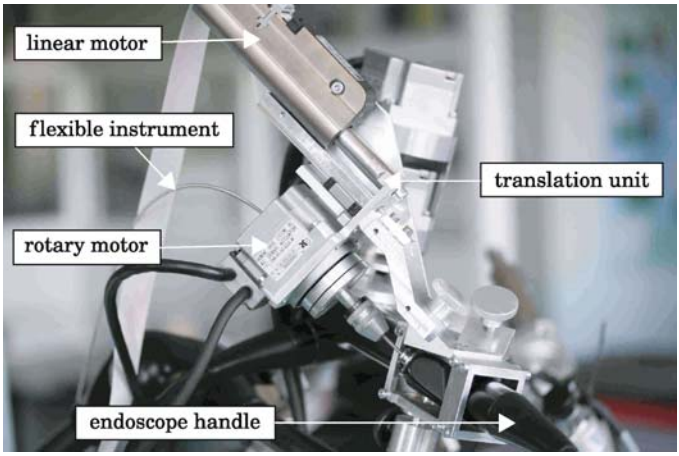


Fig. 7 Translation and rotation system of the instruments

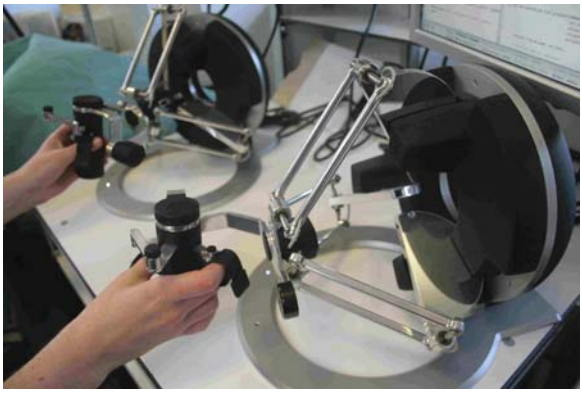


Fig. 8 The master interfaces of the prototype

which provide 7 DOFs each (shown in Fig. 8). Given the number of DOFs, it will not be intuitive for the surgeon to control each DOF separately. Several solutions are possible based on the cartesian control of the tip of the instruments, either in the endoscopic camera frame or in the base frame of the endoscope. The rotation and the translation of the instruments inside the hollow arms or the channels are decoupled from the other DOFs and can be controlled separately.

The Whole Setup

Despite the choice of the smallest and lightest equipment, the system cannot be easily carried by a surgeon. That is why a carrying cart has been designed.

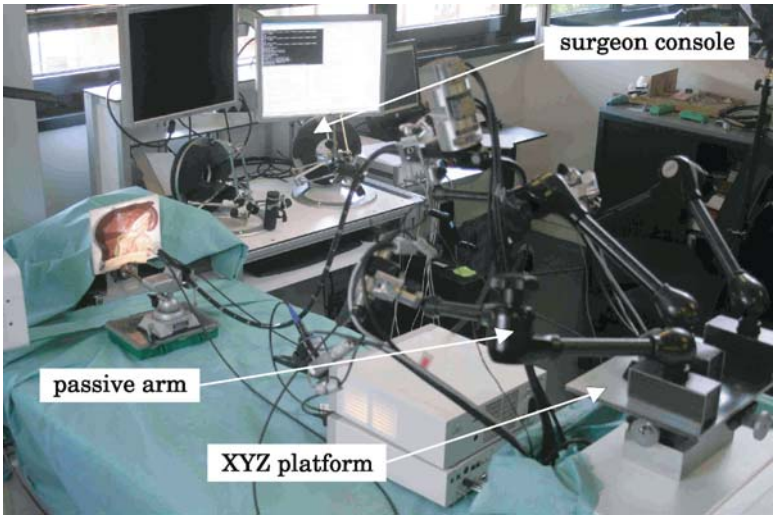


Fig. 9 View of the whole setup

This cart contains the motors amplifiers and controllers and supports a passive *XYZ* positioning table. Three passive arms which hold the motorized handles are magnetically fixed onto the platform. One arm carries the endoscope handle, its motorization, and the motorization of the instruments (translation and rotation) inside the channels. Each of the other arms carries the handle of the hollow arms and its motorization as well as the motorization (translation and rotation) of the instruments inside the hollow arms. The whole setup is illustrated by Fig. 9.

A transluminal operation using our device can be described as follows.

- First the surgeon positions the master console at the most suitable place to perform the operation.
- Then the main endoscope is inserted into the body of the patient up to the operating site. The endoscope is manually driven at the proximal part and the orientation of the bending head is controlled using a joystick.
- The cart carrying the slave system is placed at the extremity of the operating table and the handle of the endoscope is fixed on the slave console.
- The position of the base of the slave system can be adjusted with respect to the patient thanks to the *XYZ* platform or the passive arms. Then the arms are inserted into the patient's body.
- Surgical instruments can then be inserted inside the hollow arms and in the endoscope channels according to the medical requirements.
- From then on the motions of the system are controlled using the master consoles. However, if the whole system must be moved during the operation, it is possible to manually move either the *XYZ* platform, the passive arms, or even the slave cart.

The Advantages of Robotization for NOTES

The use of an intuitive master console is not sufficient. Physiological movements such as breathing motion add disturbances on the endoscopic system. The correction of these disturbances requires a complex coordination between what the hand of the surgeon controls and the motion seen in the image. That is why the endoscopic system has been robotized. Indeed, the aim of this robotic project is to teleoperate the system using master consoles while implementing automatic tasks such as visual servoing as proposed in [9]. It will also be possible to propose several working modes where surgeons could choose between dissociated motions of the arms or simultaneous use of the different elements of the endoscopic system.

Modeling of the System

The kinematic model of the system is required for providing intuitive control of the system to the surgeon. Moreover, it can be used to refine the kinematic design of the system, especially the deflection angles provided by the end cap.

Model

The bending parts of the endoscope and the arms can be assimilated to "continuum robots" since they do not exhibit any discrete joints. In [10], an interesting modeling of this kind of robot is proposed. Under the assumptions that the curvature of the continuum section is constant, that the section is inextensible and that the wires are equally spaced on the bending section circumference, a continuum robot can be modeled thanks to discrete joints, represented in Fig. 10. Thus, Denavit–Hartenberg (DH) convention can be used but interdependent DH parameters appear. This modeling can be used for the main endoscope as well as for the arms representation.

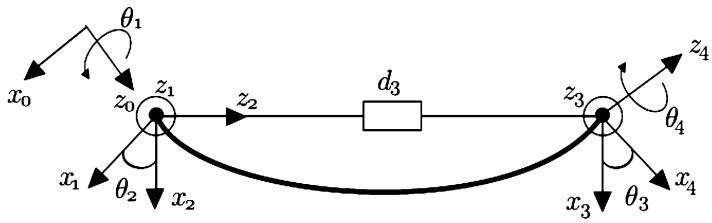


Fig. 10 Modeling of a flexible and continuum section

Table 2 DH table of a flexible section

	θ_i	d_i	α_i	a_i
0–1	θ_1	0	$-\frac{\pi}{2}$	0
1–2	θ_2	0	$\frac{\pi}{2}$	0
2–3	0	d_3	$-\frac{\pi}{2}$	0
3–4	θ_4	0	$\frac{\pi}{2}$	0
4–5	θ_5	0	0	0

The DH parameters (from Table 2) can be linked to the geometric parameters of the bending section by the following relations:

$$\theta_1 = \varphi \quad \theta_2 = \theta_4 = \theta = \frac{kL}{2} \tag{1}$$

$$d_3 = \frac{2}{k} * \sin(\theta) \tag{2}$$

$$\theta_5 = -\varphi \tag{3}$$

$$\theta_{1i} = \varphi_i \quad \theta_{2i} = \frac{k_i L_i}{2} \tag{4}$$

$$d_{3i} = \frac{2}{k_i} * \sin(\theta_{1i}) \tag{5}$$

$$\theta_{5i} = -\varphi_i \quad \text{with} \quad i \in \{B, C\} \tag{6}$$

where k is the curvature of the endoscope bending section, L the length of this section, and φ the orientation of the bending plane. k_B, φ_B, and L_B (respectively, k_c, φ_C, and L_c) refer to the same parameters for the arm B (respectively arm C). To these geometric parameters one must add I_B and φ_B (respectively I_C and θ_C) which are the translation and rotation of the instruments into the hollow arms.

The angle φ and the curvature k can be related to the length distribution of the two wires ΔL_1 and ΔL_2 by the following equations:

$$k = \frac{\sqrt{\Delta L_1^2 + \Delta L_2^2}}{Lr} \tag{7}$$

$$\varphi = a \tan 2(\Delta L_2, \Delta L_1) \tag{8}$$

where r is the endoscope radius. The arms follow the same geometric scheme:

$$k_i = \frac{\sqrt{\Delta L_{1i}^2 + \Delta L_{2i}^2}}{L_i r_i} \tag{9}$$

$$\varphi_i = a \tan 2(\Delta L_{2i}, \Delta L_{1i}) \quad \text{with} \quad i \in \{B, C\} \tag{10}$$

where r_i is the arm radius.

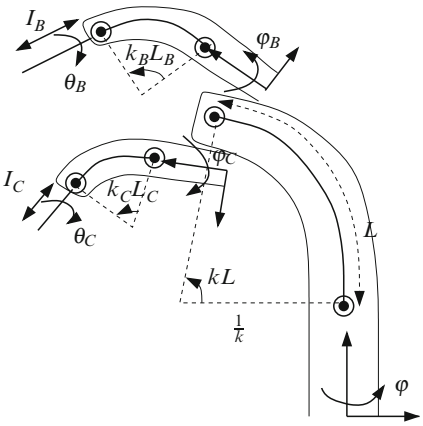


Fig. 11 Model of the head of the proposed system with 10 DOFs. Instruments in the working channels of the endoscope are not represented

The last step consists in expressing the relation between the length modification of the cables and the actuator positions $q = [q_1, q_2, q_{1B}, q_{2B}, q_{1c}, q_{2c}]^T$: with

$$\Delta L_i = R_p * \Delta q_i \tag{11}$$

$$\Delta L_{ij} = r_p * \Delta q_{ij} \tag{12}$$

with

$$i \in \{1, 2\} \quad \text{and} \quad j \in \{B, C\} \tag{13}$$

where R_p and r_p are, respectively, the radius of the endoscope pulleys and of the arms pulleys.

With this modeling, it appears that the geometric representation is singular when the length variation of the cables is null. Indeed when a flexible section is straight its curvature is null and so (2) and (5) are undefined. For this specific configuration the matrix of the geometric transformation corresponds to a translation.

The complete system is a tree-like robot with two end effectors (Fig. 11). This means that both arms have a common kinematic chain. Thus we can draw up both geometric and kinematic models. They are necessary to study the movement, the behavior and the singular configurations of the prototype, but also to optimize the kinematic design of the system.

Triangulation Analysis

The modeling is helpful to analyze the effects of arms triangulation. One of the key components of our system is the end cap which allows the deflection of the arms from the main endoscope direction. Thanks to the models, we are currently analyzing the effect of the cap angles on the working cooperation area of the arms. The objective is to optimize this area to provide a maximum workspace to the user.

For instance, one can observe in Fig. 12 that the cooperation area, A, defined by the intersection of arms workspaces (crosses on graphics), is larger when the deviation angle, ψ, is nonnull. Assuming that the arms are diametrically opposite, we have used an optimization algorithm to define the best deviation angle. For instance for $L_B = L_C = 2\,\text{cm}$, the largest cooperation area is obtained for $\psi = 26°$ (Fig. 13).

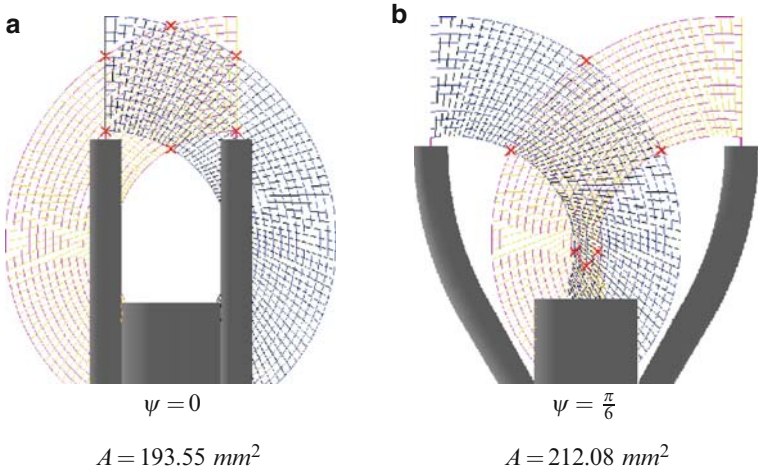


Fig. 12 Optimization of cooperation area

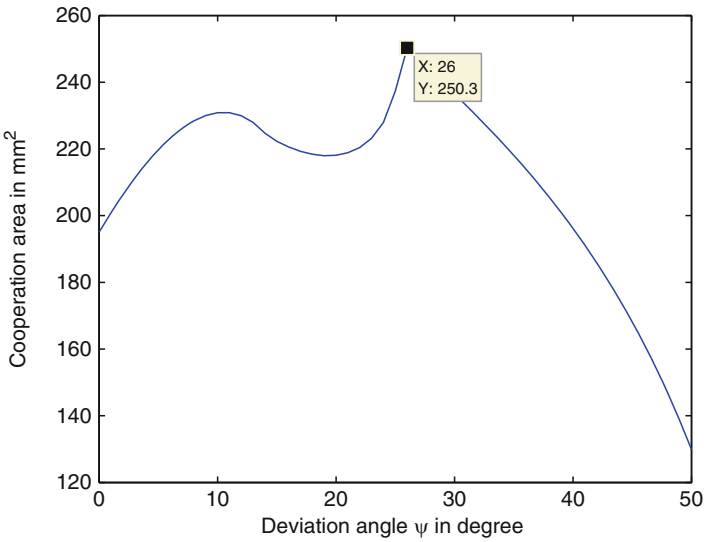


Fig. 13 Cooperation area in function of the deviation angle

Control and Telemanipulation

The system proposes two main working modes:

- A dissociated mode where surgeon can control each arm independently. The 3 DOFs of each arm can be controlled in joint space or in cartesian space according to the surgical task.
- A combined mode where surgeon can simultaneously use the different elements of the endoscopic system (endoscope and arms). Given the number of available DOFs, this mode essentially allows the surgeon to perform a complex task in cartesian space.

The telemanipulation loop from Fig. 14 fits to both modes provided that the right kinematic model is used in the computation. The reference position is given by the master interface and thanks to a numerical inversion of the geometric model previously exposed, joint positions are sent to the robot.[2]

In addition to classical control modes, a high level automatic feedback mode using visual servoing has been added to the system. The objective of this mode is to virtually link the tip of the endoscope and the anatomical target despite disturbances such as physiological motions or motions induced by the surgeon on the endoscope. The visual servoing control scheme is shown in Fig. 15. By clicking a position in the image, the surgeon selects the position s_{ref} where the area of interest must be stabilized. The position s of this area in the current image is obtained by visual tracking

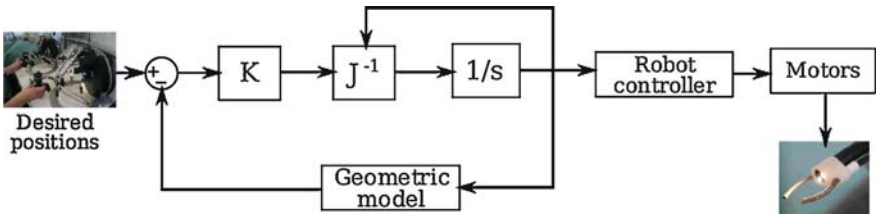


Fig. 14 Block diagram of the telemanipulation loop

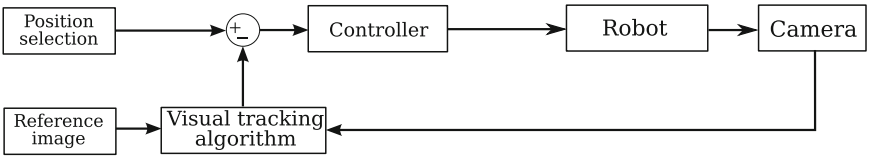


Fig. 15 Block diagram of the stabilization loop

[2] Computing jacobian matrix J results in long terms and is omitted for brevity.

[11–13]. Visual servoing techniques allow to servo to zero the error e between the reference s_{ref} and the current position s. The controller choice depends on the disturbances to reject:

- If the perturbation is unknown, the simplest controller consists of a decoupling control law involving a proportional gain. However the performance obtained by this controller in a stabilization task is not optimal. Performances can be improved by adding integral and derivative terms to the controller.
- If the perturbation is predictable (that is the case when the breathing is controlled by an external machine), a better solution is to implement a controller using a model of the disturbance such as in predictive controllers with feedback of the breathing machine signal for example [9].

Conclusion

NOTES appears as an important breakthrough for the future of noninvasive surgery. New designs of endoscopic systems is an important track for improving surgical possibilities in this field. Robotization also appears as a requirement given the high number of DOFs the surgeon has to manage. The teleoperation system we present in this chapter provides triangulation of the arms for good working conditions, and will allow several working modes coupled with automatic tasks. It has the potential to allow tasks such as stitching or knot tying. A valuable add-on could be the autonomous translation of the whole system.

Acknowledgment Authors wish to thank Karl Storz Endoskope GmbH for their support and the IRCAD for the facilities and medical advices.

References

1. Marescaux J, Dallemagne B, Perretta S, Wattiez A, Mutter D, Coumaros D (2007) Surgery without scars: report of transluminal cholecystectomy in a human being. Arch Surg 142: 823–826
2. Swanstroem LL, Khajanchee Y, Abbas MA (2008) Natural orifice transluminal endoscopic surgery: The future of gastrointestinal surgery. The Permanente Journal 12(2):42–47
3. Hattori A, Suzuki N, Suzuki S, Hayashibe M, Otake Y, Kobayashi S (2006) General development plan of surgical robotic systems. Int J CARS 1:201–228
4. Thant ZM, Low SC, Tang SW, Phee L (2006) Ergonomic master controller for flexible endoscopic gastrointestinal robot manipulator. Int. Conf. on Biomedical and Pharmaceutical Engineering
5. Can S, Fiolka A, Mayer H, Knoll A, Schneider A, Wilheilm D, Meinung A, Feussner H (2008) The machatronic support system "HVSPS" and the way to NOTES. Minimally Invasive Therapy 17(6):341–345
6. Abbott DJ, Becke C, Rothstein RI, Peine WJ (2007) Design of an endoluminal NOTES robotic system. IEEE Int. Conf. on Intelligent Robots and Systems, October, San Diego
7. http://www.olympus-global.com/en/news/2008b/nr080901endoe.cfm (2008)

8. Kobayashi T, Lemoine S, Sugawara A, Tsuchida T, Gotoda T, Oda I, Ueda H, Kakizoe T (2005) A flexible endoscopic surgical system: First report on a conceptual design of the system validated by experiments. Japanese Journal of Clinical Oncology 35(11):667–671
9. Ott L, Zanne Ph, Nageotte Fl., de Mathelin M (2008) Physiological motion rejection in flexible endoscopy using visual servoing. IEEE Int. Conf. on Robotics and Automation
10. Hannan MW, Walker ID (2003) Kinematics and the implementation of an elephant's trunk manipulator and other continuum style robots. Journal of Robotic Systems 20(2):45–63
11. Benhimane S, Malis E (2007) Homography-based 2d visual tracking and servoing. The International Journal of Robotic Research 26(7):661–676
12. Chaumette F, Hutchinson S (2006) Visual servo control, part i: Basic approaches. IEEE Robotics and Automation Magazine 13(4):82–90
13. Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(5):564–577

MRI-Guided Robot-Assisted Interventions: An Opportunity and a Challenge in Computational Surgery

Nikolaos V. Tsekos, Erol Yeniaras, and Ahmet Eren Sonmez

The Case for MR-Guided Robot-Assisted Interventions

Currently, we are witnessing the rapid evolution of minimally invasive surgeries (MIS) and image-guided interventions (IGI) for offering improved patient management and cost effectiveness. It is well recognized that sustaining and expanding this paradigm shift would require new computational methodology that integrates multimodal sensing (including imaging), controlled systems (including robots and smart actuators), the patient and the operator (e.g., [1–9] and references therein). Looking into the potential future evolution in MIS and IGI among the sought directions is the incorporation of real-time image guidance (RTIG) that can provide volumetric and high information-content visualization of the Area of Operation (AoO). Such approach would include (1) assessing in real-time tissue deformation secondary to the procedure and physiologic motion, (2) monitoring the tool(s) in 3D, and (3) updating information about the pathophysiology of the targeted tissue. With those capabilities, RTIG may facilitate a paradigm shift and methodological leap from "keyhole" visualization (i.e., endoscopy or laparoscopy) to one that uses a volumetric and information-rich perception of the AoO. This capability may eventually enable a wider range and level of complexity IGI and MIS [8, 10–12].

The "gold standard" modalities in interventional medicine are the X-ray based fluoroscopy and computer tomography (CT), and ultrasound. Compared to those modalities and in view of potential future directions in IGI and MIS, Magnetic Resonance Imaging (MRI) emerges as a strong new candidate for planning, guiding, and monitoring due to certain unique to the modality properties, reviewed in Tables 1 and 2 and discussed below in more detail. Apart from those specific properties and concomitant benefits, modern MRI scanners are endowed with a wealth of technical innovations that capitalize on the inherent physical characteristics of the phenomenon of magnetic resonance and the specific way the signal is spatially encoded to generate images. As a result, MRI is a modality that fits well as an integrated component of advanced cyber-physical systems specialized for RTIG

N.V. Tsekos (✉), E. Yeniaras, and A.E. Sonmez
Medical Robotics Laboratory, Department of Computer Science, 4800 Calhoun – 501
Phillip G. Hoffman, University of Houston, Houston, TX 77204-3010, USA
e-mail: nvtsekos@central.uh.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_10,

Table 1 From MRI-guided interventions to robot-assisted MR-guided interventions

MRI for guiding Interventions and surgeries

- Plethora of contrast mechanisms for anatomical and functional information
- True 3D and multislice 2D
- No Ionizing radiation

MR-guided robot-assisted Interventions and surgeries

- Access to the patient inside the MRI scanner (esp. high field cylindrical scanners)
- Real-time imaging for guidance and response to adjust the procedure
- Patient remains inside the MR scanner
- Generic robot features; e.g., accuracy, stability, tremor reduction

Technical challenges

- Materials suitable for a safe and effective operation in an environment with strong magnetic fields, rapidly changing magnetic field gradients and RF pulses
- Actuators that are not affected by the MR scanner and that do not affect image quality
- Effective Integration of the robotic system with the MR scanner
- Operator-centered and intuitive means of controlling the integrated system
- Means of effectively adapting and using the power of the MR modality

Table 2 Certain methodological and technical features of MRI and their application in robot-assisted interventions

Feature	Application
Multicontrast	Complementary pathophysiologic information for: • Preoperative path planning rom • Intraoperative monitoring (e.g., ablations)
Endogenous "absolute" coordinate system of the MR scanner	• Registration of the robotic manipulator • Forward kinematics calculations • Multicontrast image coregistration • Stereotactic planning
Surface RF coils	High SNR for a specific volume of interest (example, the targeted area as in Fig. 5)
True 3D or multislice 2D rom arbitrary definition of imaging planes/volumes	• Stereotactic planning based on 3D data • Automatic or semi automatic determination of access corridors: – Automatic alignment Or – Image-based manual guidance
On-the-fly adjustment of imaging planes and/or contrast mechanism	• Manipulator-driven control of imaging parameters to follow the end-effector • Freehand (manual) control of the robotic manipulator and automated forward-looking • Interactive adjustment of the acquisition strategy as needed during the procedure
Fast (subsecond) MR imaging	• On-the-fly adjustment of the procedure • Motion tracking for compensation
Miniature passive or active MR markers	• Visualization of tools in MR images • Initial registration of the robotic manipulator • MR-based calculation of spatial position for position validation and closed-loop control

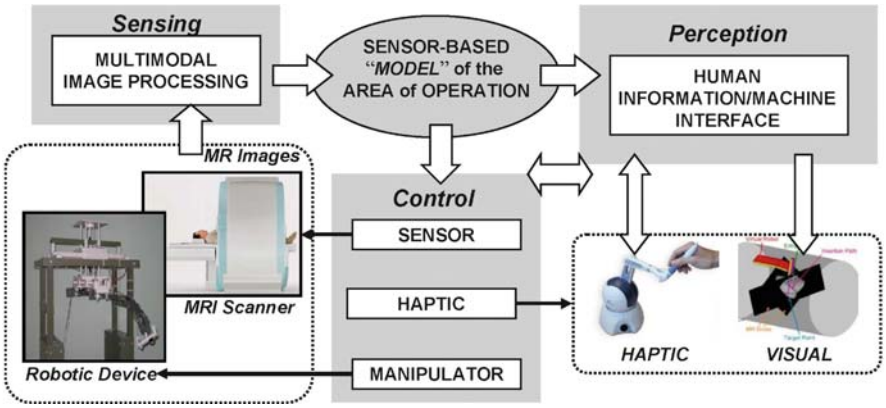


Fig. 1 Overview of a computational system for MR-guided robot-assisted interventions that integrates four interconnected elements: Sensing, Modeling, Control, and Perception. The block *arrows* illustrate the flow of data and information, and the *solid* arrows indicate the control flow. Since MRI can offer a wide range of soft-tissue contrasts that, in most cases, can be collected and updated intraoperatively, this modality offers the opportunity to use data to generate a true-"model" of the area of operation that is updated on-the-fly

procedures. Figure 1 illustrates the general concept of a system that integrates the MR scanner, the robotic manipulator, and the operator with a computational core.

As mentioned previously, MRI has certain features that are unique to the modality and make it suitable, in principle, for IGI and MIS. First, MRI offers a plethora of soft-tissue contrast mechanisms, including standard high resolution T1/ T2 contrast MR angiography, MR tissue perfusion, water diffusion, tissue thermometry, and blood oxygenation level dependent contrast. Therefore, the appropriate selection and combination of contrast mechanism, based on the pathophysiology of interest, can provide noninvasive complementary assessment of both anatomical morphology and function. As a result, MRI provides the methods for a comprehensive diagnosis and characterization of tissue pathophysiology, and the delineation of targets for biopsies and therapeutic interventions. In particular for therapeutic procedures, contrast mechanisms such as MR perfusion and thermometry allow the monitoring of contrast-altering procedures, such as thermal and cryo-ablations (e.g., [13–17]). This is exactly the practice used in diagnostic and interventional radiology, as well as in preoperative planning in certain surgical practices. With the advent of fast subsecond image acquisition schemes and enabling technologies, such as high magnetic field wide bore MR scanners and phased array radiofrequency (RF) coils, this capability of MRI can be incorporated in the IGI or MIS intra-operatively.

In addition to the rich informational content of MR data, what makes MRI, in particular a powerful modality in RTIG procedures is the flexibility and easiness for selecting the areas to be covered into images. MRI is an intrinsically three-dimensional (3D) modality that allows unrestricted and computer-controlled selection of oblique 3D or multislice imaging sets, projections, or collections of wide slabs. Figure 2 shows a representative example from an MR acquisition

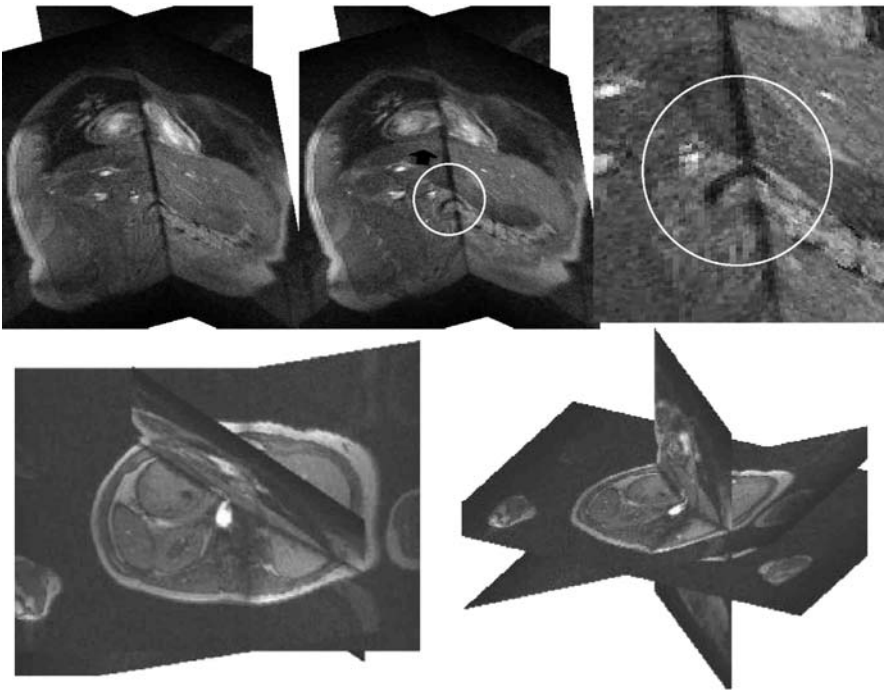


Fig. 2 In the modern MR scanners, rapid repetitive acquisition of multiple planes provide a dynamic perspective of the area of operation. The *upper panel* shows two consecutive frames from a dynamic series composed of two oblique-to-each-other slices through the abdomen of a volunteer. Note the accurate machining of the anatomical structure boundaries (especially in the zoomed-in image). The *lower panel* shows two perspectives of a frame from another dynamic MR study composed of three slices through the abdomen. Using dynamic multislice sets, the operator can monitor the robot on-the-fly, the targeted anatomy, as well as other aspects of interest within the area of operation

method, developed in our laboratory, for the dynamic collection of spatiotemporally matched and oblique-to-each-other multiple imaging planes for visualizing different anatomies and/or the interventional tools. This method is a demonstration of the flexibility of the modality, as well as the opportunity for investigators to implement data acquisition schemes that better fit the particular practice. In the case shown in Fig. 2, to visualize a procedure, specific imaging planes can be assigned and acquired, instead of a 3D volume: this allows higher rates of data acquisition and far reduced computational cost. Such operator-controlled selection of imaging planes is unique to the modality. Indeed, selection of planes is an electronic and unconstrained capability that does not require repositioning the scanner (as in CT) or manual reposition, the patient or the imaging instrument (as in the case of x-ray fluoroscopy). Moreover, modern MR scanners offer the unparalleled capability for the on-the-fly adjustment of the imaging planes, volumes, or even contrast. This is a unique modality feature that has been incorporated in vascular interventional MRI

[18–23] as well as with MR interventional robotics ([24] and following section in "Image-Guided Freehand Control").

Another property of the modality pertinent to IGI and MIS is that MRI does not use ionizing radiation, and therefore is safer for the patient and medical staff. This is of paramount importance for long imaging sessions. In addition to those three primary features, when compared with conventional ultrasound, the quality of MR images is independent of the expertise of the operator, as is the case of ultrasound. In principle, MRI provides the means for single-modality and singlesession based procedures which integrate: (1) diagnosis for identification of tissue pathophysiology and delineation of the targeted lesion, (2) guidance of the appropriate intervention, including positioning the tool and monitoring tissue property altering procedures, such as thermal ablations, and (3) assessment of the results of the interventional procedure. In view of the above benefits offered by MRI, major effort is devoted for the development and assessment of its clinical potential and merit in guiding interventional procedures [13, 25–30].

Despite the afore discussed advantages, MRI presents certain impediments in the practice of IGI: high static magnetic fields, rapidly switching magnetic field gradients, and limited access to the patient. Over the last couple of decades, new materials and actuators have been introduced most successfully to address the first two challenges. Patient accessibility however, depends on the particular design of the employed magnet. In MRI practices, the open type scanners offer the best access to the patient. With a vertical arrangement of the magnet poles, these systems offer sideways or all-around accessibility to the patient, even though the vertical distance between the poles is in the range of 40–45 cm. Although such scanners offer improved accessibility, their low magnetic fields in the range of 0.02–1.0 Tesla, result in a lower available signal and slow speed of image acquisition. In contrast, the most widely used MR scanners, and the cylindrical scanners are based on superconductive magnets that usually have a horizontal bore with typical diameters of 60 cm and lengths 1.20–2.0 m which makes access to the patient who resides at the middle of the cylinder challenging, if not impossible. Although access to the patient through the openings of cylindrical scanners, is, very limited, the cylindrical design of those scanners allow the use of magnetic field strengths of 1.0–3.0 Tesla (and most recently up to 9.4 Tesla). Such field strengths offer superior signal, speeds acquisition, image quality and magnetic field homogeneity. Prompted by the rapid evolution of interventional MR and to accommodate claustrophobic or obese patients, a new commercial system has been introduced with 70 cm diameter; this small increase in diameter does indeed make a huge difference.

As early as mid-1990s, to address this critical limitation of MRI and facilitate real-time guidance of IGI, several remotely actuated and controlled MR-compatible manipulators have been introduced [31]. With these devices, the entire or part of a procedure can be performed while the patient remains inside the MR scanner. This provides real-time visualization of the procedures and therefore, allows improved targeting while it capitalizes on the improved contrast offered by the modality. Beyond addressing the patient accessibility, manipulators offer general benefits, such as steadiness and accuracy [30, 32–36]. Currently, several examples of MR

compatible interventional robotic systems have been presented for brain biopsies [37], breast interventions [38–40], endoscope manipulation [34], prostate procedures [41–43], and general purpose systems for use with standard cylindrical MR scanners [32, 44]. Therefore, incorporation of robotic manipulator in the MR suite offers access to the patient and facilitates keeping the patient in the scanner as long as needed by the particular procedure, real-time guidance and the standard generic features of robotic manipulators.

A Developmental Platform for MR-Guided Robotic Interventions

Motivated by the concept that MRI is an information system that can be the basis of an IGI/MIS system illustrated in Fig. 1, we have implemented a prototype developmental platform MR-compatible manipulator system [44]. Viewing this as a whole, rather than as a robot, we have embarked on a systems approach to develop and investigate different enabling-technologies and approaches for performing robot-assisted interventions with MR guidance [24, 44–49]. Progressively, this system has been modified to develop and demonstrate different MR-based enabling technologies listed in Table 2. Figure 3 illustrates the architecture and the

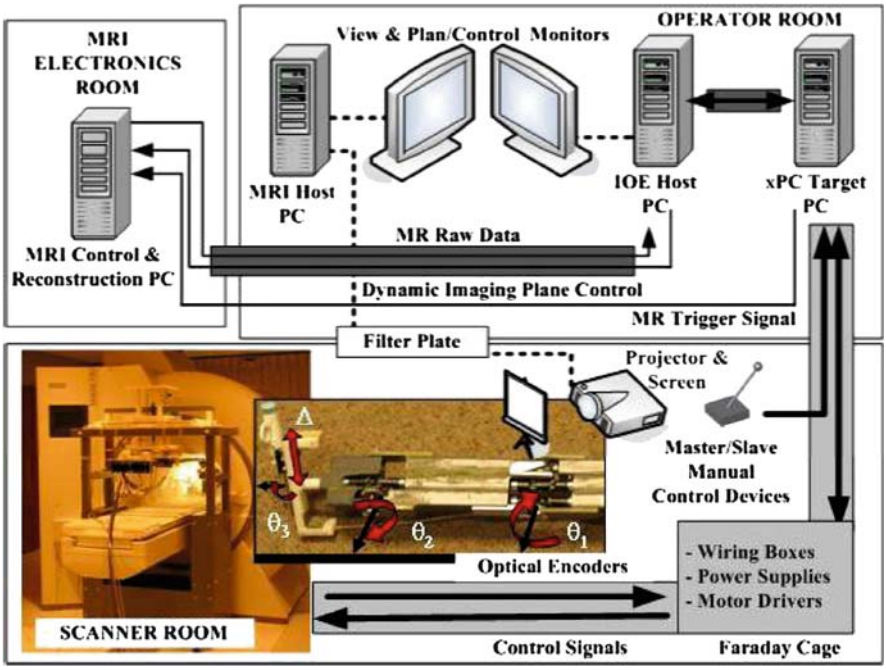


Fig. 3 The architecture of the prototype robotic system for MR guided interventions. Figure adopted from [47]

primary components of this system: an MR compatible robotic manipulator, its associated control hardware and software components, human–machine interfacing and an MRI scanner interconnected through a computational core component.

The manipulator is a seven degree-of-freedom (DOF) robot that consists of two parts: a base and an articulated arm [44,49]. The base is located in front of the opening of the gantry and is equipped with a three orthogonal DOF Cartesian positioner that carries and maneuvers the arm along the three principal axes (X, Y, and Z) of the MR scanner coordinate system. The arm, which extends inside the gantry of the scanner and above the patient, has three computer-controlled rotational and one manual linear DOF. Two of these rotational DOF (θ_1 and θ_2) resemble a dual "elbow" intandem and used either individually or in accord to set the first Euler angle (ϕ) on the vertical plane (i.e., the sagittal plane YZ of the scanner). The third rotational DOF (θ_3) resembles a "wrist," which is orthogonal to the axis of θ_2, and is used to set the second Euler angle (θ). With the combination of those six computer controlled DOF, the end-effector can be placed at any desired position and orientation within the area of its operation. The end-effector of this manipulator is equipped with a carriage on which an interventional tool, such as a biopsy needle or an RF ablator, can be attached. A remotely actuated manual trigger mechanism allows the release of the carried tool at will. The carriage can be translated by means of a manually actuated cable-driven mechanism for advancing the carried tool; this is the seventh and linear DOF of the manipulator.

Actuation of the six computer-controlled DOF is achieved with commercially available ultrasonic rotary motors [31]. To optimize MR compatibility and allow the implementation of an arm with a slim profile, the ultrasonic motors that actuate the three computer-controlled DOF of the arm are located outside the gantry, and motion is transferred through joint articulated transmissions [49]. To eliminate noise on the MR images, all electronic power supplies are housed in a Faraday cage while all cables are shielded with several layers of aluminum sheath. In vitro and in vivo studies have demonstrated that this shielding is efficient for real-time MRI guidance even when the robotic device was actuated during image collection [44,47,48].

The hardware architecture of the system, shown in Fig. 3, was primarily dictated by our choice to develop the computational core of the system on Matlab and the Simulink (The Mathworks, Inc., Natick, MA) based xPC Target real-time environment. This choice was proved an efficient and straightforward way for rapid prototyping, allowing us to focus on the methodological aspect and investigating new ideas in image-based control without extensive distractions, e.g., human power and time-intensive software development. As illustrated in Fig. 3, the system utilizes two dedicated personal computers (PC), the "Host PC" and the "Target PC." Overall, the hardware of the system is composed of five distinct components. (1) The "Host PC" is physically connected to the MR scanner and the "Target PC" via dedicated Ethernet cards. The "Host PC" runs the core control software and provides means for human–machine interfacing through a graphical user interface (GUI), a joystick and a master/slave device [47]. From the MR scanner, raw MR data that reconstruct and present them on dedicated windows on the GUI. The system is capable of presenting and manipulating oblique sets of dynamic data like these shown in Fig. 2.

The control core generates and sends appropriate instructions to the "Target PC," for real-time control of the manipulator, and to the MR scanner, for on-the-fly adjustment of the position and orientation of the imaging plane(s). (2) The "Target PC" runs the real-time xPC Target that also houses an array of cards, such as digital input/output, counters, digital-to-analog converters for interfacing to the ultrasonic motor controllers, optical encoders on the motors and the master/slave control device, and optionally generating triggering TTL pulses to the MR scanner. (3) The Master/Slave Manual Control device offers on-the-fly freehand control of the robot and, via the target-to-host communication, control of the position, and orientation of the MR imaging planes. (4) The robotic manipulator described above. (5) The control electronics.

The functional structure of the computational core resembles the one shown in Fig. 1 and integrates: (1) sensing, i.e., the MR scanner, (2) modeling of the AoO by means of hierarchically organized preoperative and intraoperative raw or processed MR images, (3) control that manages both the six actuated DOF of the manipulator and the MR scanner, and (4) perception and human–machine interfacing. At its current state, compared to the system described in Fig. 1, this prototype developmental platform lacks a haptic interface and the "sensor based model of the AoO" is at an early stage. This computational core consists of several components that provide the needed tools for operator-supervised (manual) functions and performs several tasks autonomously:

- *Image component*: This component interfaces to the MR scanner to receive data, processes them and directs them to the other components of the computational core. Via a dedicated TCP/IP connection to the local network of the scanner, this component uploads to the "Host PC" MR data in two ways: either as operator-selected previously collected or automatically during real-time imaging. In either case, after transferring the data to the "Host PC," this component can reconstruct raw or DICOM format images. Appropriate routine extract from the header of the data file-specific information, such as the position and orientation of the imaged plane, the acquisition matrix size, the field of view, and the type of contrast (i.e., pulse sequence). In the case of raw data, the code also extracts information about the specific acquisition scheme, such as segmentation, acquisition matrix, and order of segment acquisition, in order to reorder the k-space lines and reconstruct the images with Fast Fourier Transform. This component reconstructs multislice sets allowing the visualization of the movement of the robot with a 3D or volumetric perspective. Figure 4 illustrates an example of monitoring the maneuvering of the robotic manipulator using four planes (collected with the same method as shown in Fig. 2). The component is expanded to include image processing algorithms, such as the segmentation of blood vessels from time-of-flight MR brain angiography and generation of 3D binary masks for automatic planning. After reconstruction, and processing, images or processed masks are stored locally at the "Host PC." A log-array hierarchically records the type of contrast or equivalently the processed mask with an identifier flag. An example of such processed data is the 3D surface used in the safety component discussed below. When a component of the computation needs certain type of contrast or

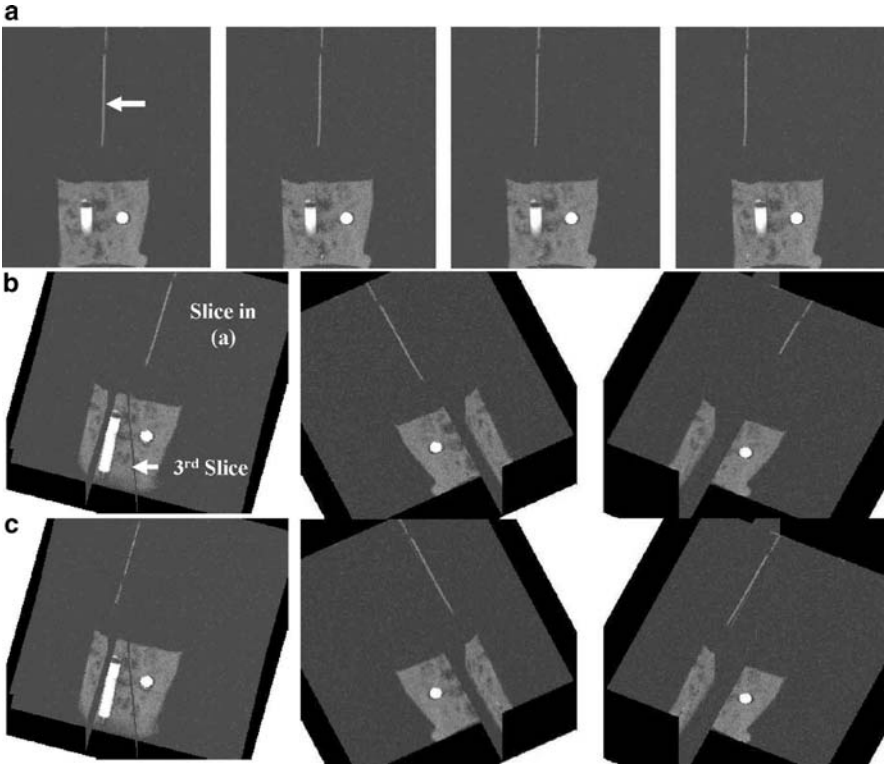


Fig. 4 MR-guided stereotactic control of the robotic manipulator described in [44] collecting multislice sets with the same interleaved method is shown in Fig. 2. In these studies, a phantom was used with two tubes filled with Gd-based contrast agent embedded in a matrix of fat. The needle was also filled with contrast agent allowing its clear visualization on those T1-weighted images (*white arrow*). The panel in (**a**) shows four frames of a transverse slice (out of four used in this multislice set) collected during the actuation of one of the linear DOF along the x-axis of the MR scanner. (**b**) and (**c**) show the 3D reconstruction of two planes (out of four) in pairs in the beginning (**b**) and at the end of the motion (**c**). The three pairs correspond to three different perspective views of the reconstruction

processed information, it refers to this array and screens for the appropriate identifier flags to select and access those specific information from the data base.

- *Kinematics component*: This component is an assembly of routines that are collected by other components of the computational core on a per-need-basis to calculate forward and inverse kinematics of the manipulator. All those calculations are performed relative to the coordinate system of the MR scanner. The initial position of the manipulator is measured from the registration component described below.
- *Registration component*: This component performs the registration of the manipulator to the coordinate system of the MR scanner [44, 45]. This registration is necessary for determining the initial position of the manipulator and

the original set of the DOF values to be used in the forward and inverse kinematic calculations. For registration, the manipulator is equipped with a set of 3D cross-shaped markers made of three orthogonal to each other tubes filled with diluted Gd-based contrast agent. Those markers are attached to specific positions on arm and the end-effector. After the placement of the robotic arm inside the scanner, a multislice set of T1-weighted scout images is collected and from them are extracted the coordinates of the centers of all the markers. Since their position on the arm and end-effector is known, the initial position and arrangement of the manipulator relative to the coordinate system of the MR scanner is determined. This information is then used by the forward and inverse kinematics component. This routine may be run only once at the beginning of a study, or it can be repeated by the operator if for any reason the reset of the system is desired. An additional means is also used for validating the position of the robot after maneuvering (i.e., from the position of the markers we extract the values of the DOF).

- *Safety component*: The computational core is endowed with an autonomously running and continuously active safety routine that prevents any moving part of the manipulator from colliding with the subject or the gantry of the scanner [44, 47]. This routine has an authority over any process that may alter the configuration of the manipulator; any actuating instruction, even a direct one from the operator, is first validated by this routine and then is performed. A software (with a GUI button) and a hardware (a switch) override is allowed. This is an image-based function: After the initial registration, a transverse multislice set is collected and the edge of the subject is extracted and interpolated to a 3D surface in the coordinate system of the MR scanner. The space between this surface and a cylinder that has the same diameter with the gantry of the specific scanner, defines the volume of allowable presence of the end-effector and any point of the manipulator. Notably, the inherent coordinate system of the MR scanner makes those calculations straightforward and computationally fast. In the latest version of our protocols is combined the performance of registration, with this one to share the same MR data (to further speed up setting up the system).
- *Scanner control component*: A dedicated component of the computational core directly controls the MR scanner via a dedicated TCP/IP protocol and Ethernet card on the "Host PC" [47]. When activated, this routine uses as input the same control instructions sent to the manipulator to calculate from the forward kinematics the anticipated position of the end-effector relative to the coordinate system of the MR scanner. After conversion to appropriate values, send them to the MR scanner for on-the-fly adjustment of the orientation of the imaging planes [47]. In case that triggering is desired for the collection of MR data, this routine accesses one of the digital I/O cards on the xPC to generate the appropriate single pulse or sequence of TTL signals. This approach has been proved useful for the intermittent collection of images after an action or when image collection is desired immediately after an actuation step rather than during robot actuation.
- *Manipulator control component*: This component manages the control of the manipulator and offers stereotactic guidance with autonomous control and free-hand guidance with master/slave control, and will be discussed in detail in the following two sections.

This manipulator has been used for developing different means of MR image-based control and performing an array of studies, including stereotactic [24, 44–46], free-hand [45, 47] and tele- [48] MR-guided interventions. Implemented on a scalable and modular way, it can be easily expanded to include additional functionality or to implement different paths of data flow and sharing among the different components for autonomous or assisted-to-the-operator decision making.

Image-Guided Stereotactic Control

Stereotactic guidance and (semi) autonomous control of the robotic manipulator in one form or the other is used in the majority of the current MR compatible systems. Based on the 3D and multicontrast nature of the MR modality, stereotactic control is an excellent demonstration of power and benefits of combining MR guidance and robotic guidance. With this approach the operator inspects MR images to prescribe a path for the insertion of the interventional tool to a targeted tissue. It should be emphasized that with MRI collection, those images may occur while the patient and the robot are inside the MR scanner (i.e., just prior to the procedure). Since the robotic manipulator is registered to the MR scanner, and the patient does not change position (or transferred between the imaging and stereotactic suite), this eliminates the need to (re) register the images and/or the patient to the stereotactic system. Using some type of graphical marking on the MR images, the operator defines the targeted tissue (e.g., a lesion) and a point of insertion on the external surface of the patient (i.e., on the skin). Due to the 3D nature of MRI and the capability to collect images with any oblique orientation, the two points can be selected to two different planes not even parallel to each other. The operator may also use images of different contrast to assess different anatomical and pathological features thereby defining the most appropriate approach to the targeted anatomy. This path of access maybe selected initially by the operator manually, and then further adjusted by routines to ensure that this path of access does not include blood vessels or other vital tissues.

The prototype system described herein offers the means to perform stereotactic guidance as described and demonstrated previously [24, 44, 46]. In addition to the aforementioned planning steps, a newer version of the stereotactic control of the system incorporates two additional layers of autonomous processing that run and advice the operator during the selection of the access path: One imposes and resolves kinematics and safety constraints, and the other patient-specific anatomy constraints. The first layer uses the three above described kinematics, registration and safety components to impose constraints on the selection of the path. Specifically, it continuously calculates the reverse kinematics for any path and evaluates whether it is within the operational envelop of the kinematic structure of the manipulator. In practice, any adjustment of the insertion or target point triggers the inverse kinematics and safety components. If this new adjustment is not compliant with both the kinematic constraints (i.e., whether the manipulator can deliver it) and the safety limitations (i.e., whether a kinematically allowable movement does not

infringe to safety), the operator is warned and the selection is canceled. The second layer provides the operator with an additional constraint based on anatomical features. Specifically, the code uses the selected points of the insertion path to define a vector in the coordinate system of the scanner and compares this array with any mask stored in the log-array. If the vector intercepts, a specific anatomical feature (e.g., a vessel in an angiography mask) is also rejected. This particular component has been proved a time-efficient method for inspecting the selected path relative to the tissue in between the point of insertion and the point on the target.

Upon approval of a path by the operator, the control component of the core software actuates the DOF of the manipulator to position the end-effector in a way that the axis of the interventional tool is aligned with the selected path. The particular sequence of actuation is based on a simple hierarchical order: First the Z Cartesian DOF (along the axial dimension of the long axis of the cylindrical MR scanner), then the X and Y are actuated simultaneously followed by the sequential actuation of the three rotational DOF: θ_1, θ_2 and θ_3. It should be noted that for this particular sequence of actuations, the safety routine has already checked the appropriateness of each step; therefore, ensuring that no collision may occur for any intermediate transient position of the manipulator. It should be noted that, this approach is suitable in regard to safety during the actuation of the robot since the safety component can use end-inspiration and end-expiration 3D surfaces of a subject. However, the method is appropriate to achieve accurate target alignment only with nonmoving tissues, as example in neurosurgical procedures or abdominal interventions combined with breath holding.

Image-Guided Freehand Control

Stereotactic guidance combined with the anatomical and functional information offered by multi contrast MRI is a powerful tool for a plethora of interventions. However, recent technological innovations available on the state-of-the-art MRI scanners offer some intriguing opportunities: Performing procedures with the operator manually controlling on-the-fly both the robotic manipulator and the MR scanner in accord. Such an integration of the interventional manipulator with the imaging modality maybe appreciated if we consider two otherwise unrelated practices: the standard ultrasound guided interventions, such as the routinely performed biopsies, and the robotic surgeries like those performed with the da Vinci Surgical System (Intuitive Surgical, Inc. Sunnyvale, CA). In both practices, we may identify two common elements: Man-in-the-loop control (i.e., the radiologist who holds and maneuvers both the ultrasound probe and the biopsy needle or the surgeon who controls the robot) and real-time forward looking appreciation of the area of interest (i.e., the ultrasound probe and the video camera on the da Vinci robot). In both cases, there is an immediate involvement of the operator who relays on real-time updated visual information. Modern MRI scanner can offer this capability by means of interactively adjusting the imaging planes thereby resembling the practical aspects

of the above mentioned practices: The movement of the ultrasound probe to scan the area and the movement of the endoscopic camera of the da Vinci System that is attached on the robotic arm.

In the case of MRI, a manually or "freehand" controlled robotic manipulator can be interfaced to and take advantage of the capabilities of the modern state-of-the-art MRI scanners that allow on-the-fly adjustment of imaging parameters, including the position and orientation of imaging slices. This capability is based on the unique modality property of setting the acquisition parameters in a straightforward computer-control manner. For example, the orientation of the imaging plane in MRI is determined by the orientation of a magnetic field gradient that is applied during the process of slice selection by changing the relative combination of three orthogonal magnetic field gradients, a slice with any oblique orientation relative to the coordinate system of the scanner can be chosen. Such adjustments could be done before commencing image acquisition, but recent technical advances allow changing such acquisition parameters on-the-fly. In practice, the MR scanner is set to continuously acquire a slice, and during the course of this acquisition the operator can manually (or a routine automatically) change the acquisition parameters that are applied to the immediately following frame. An external source of updated parameters can also be used to feed the controller of the scanner. This capability was initially demonstrated in vascular MR-guided interventions were the transient position of a catheter, measured with appropriate MR methods, was used for on-the-fly adjustment of MRI imaging parameters to follow the movement of catheter dynamically [18–23].

Recently, Christoforou et al. demonstrated the integration of the prototype MR compatible robotic manipulator discussed herein with a commercial MRI scanner [47]. Using the aforediscussed Scanner Control component, the transient position of the robot was sent to the controller of the MR scanner, which adjusted the position and orientation of the imaging plane to visualize the end-effector of the system. This mode allows the operator to maneuver the end-effector above the area of interest while scanning the anatomy as shown in Fig. 5. The ability for man-in-the-loop direct control of the interventional tool, combined with the on-the-fly update of the imaging plane, corroborates to a simple and intuitive image guidance in a way similar to this with ultrasound-guided interventions, image-guided interventions, or with endoscopy-based surgical robots. The manipulator-driven real-time imaging provides for scouting the subject, identifying a target, and setting the path of the interventional tool to clear of obstacles and align it to the target. Man-in-the-loop image-guided control may also provide the means of practice for compensation of needle bending, a major source of error observed in previous studies with MR compatible systems error [33,42,43,50], since the operator can use dynamic imaging to react appropriately and to correct the bending.

The feature of having the tool always at the same position and orientation relative to the FOV provides a straight-forward way of directing the tool, while a simple software routine can place a line-of-sight on any frame without any special image processing. Figure 6 illustrates an example on how to control the robot manually using real-time MR images with simple graphical objects overlaid to assist manual control with visual input. In this particular case the operator has set a corridor of

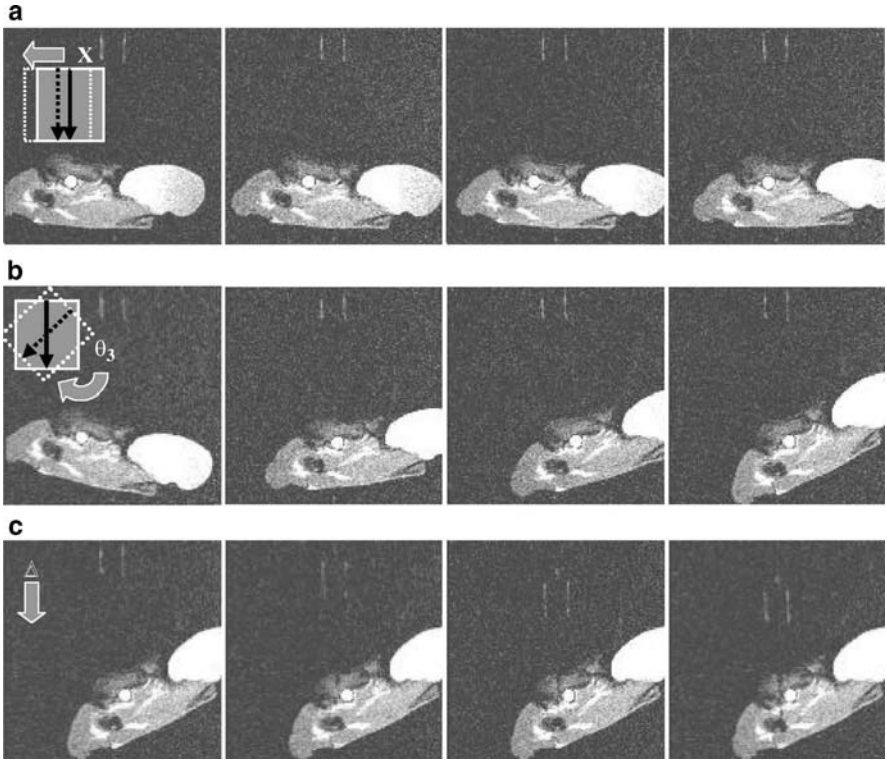


Fig. 5 Selected frames from a freehand robot control during the actuation of the (**a**) linear X, (**b**) rotational θ_3, and (**c**) linear insertion DOF of the manipulator. The position and orientation of the imaged slice is updated dynamically based on the position and orientation of the end-effector as calculated by the control software, so the latter is always seen at the same position relative to the FOV and parallel to the gradient axes. Adopted from ref. [47]

access (two parallel lines) to the targeted anatomy (high intensity structure). This corridor was defined from an original set of multislice scout images. During the maneuvering of the robot, the computational core sets a line-of-sight to assist the operator directing the end-effector. For generating the line-of-sight, two approaches have been implemented. The first is kinematics-based: Since the transient configuration of the manipulator is known relative to the MR scanner coordinate system, from the initial registration and the forward kinematics such a line can be generated by the core software and superimposed on each frame. The second approach is image-based: The MR images are processed by the core software on the "Host PC" to identify the location of the two MR markers on the end-effector, and then superimpose a line in between them (i.e., where the interventional tool is located). The coordinates of the MR markers (are extracted from the following sequence of tasks: Median filtering to reduce the noise, Otsu thresholding for binary segmentation, morphological skeletonization, and Hough transform. The coordinates of the two

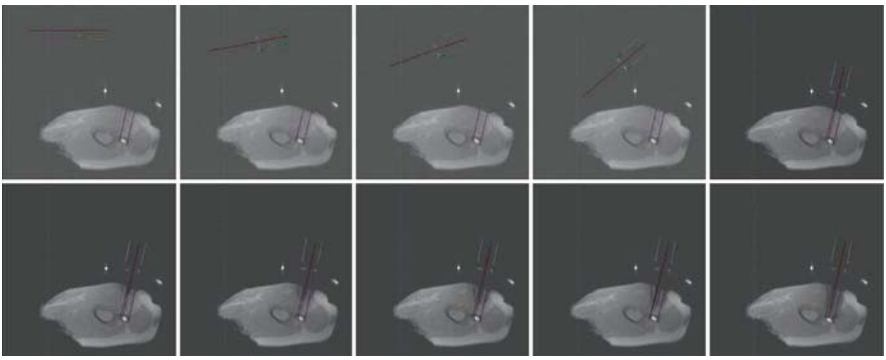


Fig. 6 Demonstration of computer-assisted robot-mediated MR-guided simulation of an intervention with the robotic device described in [44]. In this study, the phantom was a piece of meat with a Gd-filled bottle to simulate a contrast enhanced target. A corridor of access (two *parallel lines*) was prescribed from analysis of multislice scout images. The operator controlled the robot using the manual device described in [47]. As the operator manually controls the robot, the control software performs the following tasks seamlessly (pertinent to Fig. 1): It sets commands to the MR scanner for on-the-fly change of the imaging planes based on the actuated DOF, triggers the collection of the MR images. Knowing the position of the end effector, the GUI places a graphical line along the line-of-sight to assist the operator

markers were transformed to the coordinate system of an OpenGL screen by a Custom Affine Transformation, and the line is designed by the appropriate OpenGL object. The access corridor was also calculated and drawn for each frame. In either case, the line-of-sight proved a very simple and efficient way for visually guiding the manipulator using real-time MR images.

With this system, manually control of the manipulator is performed with a simple master/slave control device [45,47]. This device is composed of two-pieces each and is used for the control of specific DOF: one piece is an articulated handle, which resembles the kinematic structure of the arm with three joints corresponding to the three rotational DOF, and the other piece has three dials, which correspond to the three Cartesian DOF of the base. The joints and the dials are connected to optical encoders to monitor their changes by the operator. Those readings are converted to motion instructions on the "Target PC" and used as input to the control component. Operator-selected amplification parameters are used to set the sensitivity of the response of the robot. Although the master/slave control device used with the prototype system presented herein has been very instrumental in the development of the discussed "manipulator-driven" imaging, is of a preliminary version and lacks certain features. Our studies clearly demonstrates that such a device must incorporate active actuation, so that the operator can sense important information, such as backlash or motion constraints, imposed by the safety check component. Future work will address this limitation developing a haptic interface for the manual control of the robotic manipulator. While such a master/slave approach is common in robotics, it is the combination with the MRI scanner that endows the system for an innovative image-guided master/slave control scheme.

Future Perspectives

The inherent modality-unique properties and a wide range of enabling technologies, reviewed in Tables 1 and 2, have advanced the field of interventional MRI. Innovative works in robotic manipulators have demonstrated both the practice potential of those devices as well as the resourcefulness of means to further advance the field. From the technological and methodological points of view, MRI is a very versatile and powerful imaging modality, and when combined with robotic manipulators may improve current practices and open new horizons in interventional medicine. However, it is the clinical merit of this exciting technology that is of paramount importance and will determine its future evolution and fate eventually [45]. Several nontechnology-related factors should be considered when we look into the future of this field and the two questions, we should continuously ask ourselves are: "Do we really need it," and "How much will it cost."

The first and probably most critical issue is whether MRI-guided robot-assisted procedures do offer substantial benefits to patient management as compared with current interventional practices, and/or whether this technology may offer the means to facilitate new procedures, that are not possible with other modalities. Today, MRI is a tool used either for pre-operative planning or as guiding procedures by interventional radiologists, cardiologists, surgeons, orthopedic surgeons, and neurosurgeons. At the research level, new paradigms have also immerged even including cardiac valve replacement under MR guidance [51]. However, the transition to clinical practice would require substantially increasing the number of clinically viable applications and trials. Current studies are rather limited, and MR compatible interventional robotic systems have used in clinical studies on the breast [33, 50], spine [52], and prostate, [43]. Within this framework, an important factor is also the long term benefit to the individual patient in regard to quality of life. While overall MIS and IGI do result in improved patient quality of life, the same criterion is relevant when contemplating MR-guided robot-assisted procedures for example, transitioning from a well established laparoscopic to an MR-guided robot-assisted procedure. Thus, when designing new systems first we should always consider the end-user and the end-beneficiary, and then "reverse engineer" the technological solution.

Assuming that clinically viable and meritorious applications are identified, the next important factor is cost. Especially in an era of limited financial resources for research and development, managed healthcare, and ever-growing number of cost-conscious patients and healthcare entities, cost plays an ever-growing role. We are all witnessing the intense scrutiny of high-technology systems based on mere financial criteria. The financial aspect includes several interrelated factors: the cost of acquiring a robotic system, operation and consumables, space allocation, training of the personnel, maintenance and upgrades of the devices, and most importantly, mechanisms of reimbursement. In particular, with MR guided interventions, one should also consider that MRI is one of the most expensive modalities to operate. As a consequence, even the duration of setting up the system for performing the procedure may eventually be important and may need to be considered in the designs of such systems.

Assuming that clinically meritorious and financially viable practices are identified, then the power of combining MRI guidance and robotic assistance can be further exploited. The work of many groups worldwide has endowed technological armamentarium with able methodology to pursue this emerging tool (some of them are reviewed in Table 2). Looking into the future, we believe that of paramount importance is the effective integration of MR imaging with the operation and control of the manipulator in order to harness the capabilities of the modality in simple, intuitive, and effective ways. One example of such integration has been demonstrated by our group that implemented a manipulator-driven MR guidance ([47]; Figs. 3 and 5). A similar concept has also been tested in accord with stereotactic guidance [46]. Those preliminary studies indicate that both methods are benefited from controlling the MR scanner the control software of the manipulator taking advantage of the 3D and multislice capabilities of the modality. Current work focuses on incorporating preoperative high resolution and high contrast images.

Our preliminary experience from those studies supported the notion that often too much information may not necessarily serve operator [31, 45–47, 53]. Three-dimensional and multi-contrast information may increase the work load, distract the operator, and lengthen the procedure. Depending on the physician and the procedure, a simple freehand approach like this reported by Christoforou et al. [47] may be far more intuitive since it resembles the performance of a procedure with the gold standard ultrasound imaging. Indeed, pilot studies with the "manipulator-driven" on-the-fly control the MR scanner proved an efficient way for maneuvering the manipulator, by offering "forward-looking" capabilities and at-will scouting of the target [47, 53, 54]. Previous studies have illustrated that MRI with different contrast can be segmented and overlaid to define pathways and avoid sensitive structures (such as vessels) and healthy tissues [25, 27, 28, 55]. Always having the end-user as the primary determinant of a design, an additional aspect that needs careful evaluation is the human–machine interface.

Based on aforementioned observations, as well as anecdotal discussions with colleagues, we have reached the conclusion that it is important to focus on research and development on the MR scanner and an efficient computational core that integrates all components to an information system. In the ideal system, instead of directing a cascade of data to the operator (multicontrast and 3D imaging, patient condition, procedure planning and dynamic update, and control of the manipulator), such a system will provide the operator with decision-making quality information. Maturing those multimodal and multidimensional data to meaningful information is a critical aspect in the operating room of the future (irrespective of this is an interventional radiology or cardiology practice or a surgical suite). We advocate that such a computational core should be able to manipulate and customize on-the-fly raw MR data acquisition. Endowed with event-based reactivity, this controller will also adjust on-the-fly the MRI scanner to supplement missing information.

Acknowledgments NVT is grateful to Professors Eftychios Christoforou (University of Cyprus), Alpay Ozcan and Christine Menias (Washington University) for their contributions. This work was supported in part by NSF CPS-0932272, HEAF 97722, and NIH RO1HL067924.

References

1. Anderson CA, Kypson AP, Chitwood WR Jr (2008) Robotic mitral surgery: current and future roles. Curr Opin Cardiol 23:117–120
2. Anderson CA, Rodriguez E, Chitwood WR Jr (2007) Robotically assisted coronary surgery: what is the future? Curr Opin Cardiol 22:541–544
3. Boyd WD, Kodera K, Stahl KD, Rayman R (2002) Current status and future directions in computer-enhanced video- and robotic-assisted coronary bypass surgery. Semin Thorac Cardiovasc Surg 14:101–109
4. Cohn LH (2006) Future directions in cardiac surgery. Am Heart Hosp J 4:174–178
5. Goh P, Tekant Y, Krishnan SM (1993) Future developments in high-technology abdominal surgery: ultrasound, stereo imaging, robotics. Baillieres Clin Gastroenterol 7:961–987
6. Menciassi A, Quirini M, Dario P (2007) Microrobotics for future gastrointestinal endoscopy. Minim Invasive Ther Allied Technol 16:91–100
7. Nathoo N, Vogelbaum M, Barnett G (2005) In touch with robotics: neurosurgery for the future. Neurosurgery 56:421–433
8. Reijnen MM, Zeebregts CJ, Meijerink WJ (2005) Future of operating rooms. Surg Technol Int 14:21–27
9. Taylor GW Jayne DG (2007) Robotic applications in abdominal surgery: their limitations and future developments. Int J Med Robot 3:3–9
10. Woo YJ (2006) Robotic cardiac surgery. Int J Med Robot 2:225–232
11. Muntener M, Ursu D, Patriciu A, Petrisor D, Stoianovici D (2006) Robotic prostate surgery. Expert Rev Med Devices 3:575–584
12. Ito F Gould JC (2006) Robotic foregut surgery. Int J Med Robot 2:287–292
13. Aschoff AJ, Rafie N, Jesberger JA, Duerk JL, Lewin JS (2000) Thermal lesion conspicuity following interstitial radiofrequency thermal tumor ablation in humans: a comparison of STIR, turbo spin-echo T2-weighted, and contrast-enhanced T1-weighted MR images at 0.2 T. J Magn Reson Imaging 12:584–589
14. Boaz TL, Lewin JS, Chung YC, Duerk JL, Clampitt ME, Haaga JR (1998) MR monitoring of MR-guided radiofrequency thermal ablation of normal liver in an animal model. J Magn Reson Imaging 8:64–69
15. McDannold NJ Jolesz FA (2000) Magnetic resonance image-guided thermal ablations. Top Magn Reson Imaging 11:191–202
16. Merkle EM, Shonk JR, Duerk JL, Jacobs GH, Lewin JS (1999) MR-guided RF thermal ablation of the kidney in a porcine model. AJR Am J Roentgenol 173:645–651
17. Tacke J, Speetzen R, Adam G, Sellhaus B, Glowinski A, Heschel I, Schaffter T, Schorn R, Grosskortenhaus S, Rau G, Gunther RW (2001) Experimental MR imaging-guided interstitial cryotherapy of the brain. AJNR Am J Neuroradiol 22:431–440
18. Atalar E, Kraitchman DL, Carkhuff B, Lesho J, Ocali O, Solaiyappan M, Guttman MA, Charles HK Jr (1998) Catheter-tracking FOV MR fluoroscopy. Magn Reson Med 40:865–872
19. Hillenbrand CM, Elgort DR, Wong EY, Reykowski A, Wacker FK, Lewin JS, Duerk JL (2004) Active device tracking and high-resolution intravascular MRI using a novel catheter-based, opposed-solenoid phased array coil. Magn Reson Med 51:668–675
20. Zhang Q, Wendt M, Aschoff AJ, Zheng L, Lewin JS, Duerk JL (2000) Active MR guidance of interventional devices with target-navigation. Magn Reson Med 44:56–65
21. Zhang Q, Wendt M, Aschoff AJ, Lewin JS, Duerk JL (2001) A multielement RF coil for MRI guidance of interventional devices. J Magn Reson Imaging 14:56–62
22. Karmarkar PV, Kraitchman DL, Izbudak I, Hofmann LV, Amado LC, Fritzges D, Young R, Pittenger M, Bulte JW, Atalar E (2004) MR-trackable intramyocardial injection catheter. Magn Reson Med 51:1163–1172
23. Wacker FK, Hillenbrand CM, Duerk JL, Lewin JS (2005) MR-guided endovascular interventions: device visualization, tracking, navigation, clinical applications, and safety aspects. Magn Reson Imaging Clin N Am 13:431–439

24. Ozcan A, Christoforou E, Brown D, Tsekos N (2006) Fast and efficient radiological interventions via a graphical user interface commanded magnetic resonance compatible robotic device. Conf Proc IEEE Eng Med Biol Soc 1:1762–1767
25. Debatin JF, Adam G (1998) Interventional magnetic resonance imaging. Springer, Berlin
26. Jolesz F, Kahn T, Lufkin R (1998) Genesis of interventional MRI. J Magn Reson Imaging 8:2
27. Jolesz FA (1998) Interventional and intraoperative MRI: a general overview of the field. J Magn Reson Imaging 8:3–7
28. Lufkin RB (1999) Interventional MRI. Mosby, St. Louis
29. Ladd ME, Zimmermann GG, Quick HH, Debatin JF, Boesiger P, von Schulthess GK, McKinnon GC (1998) Active MR visualization of a vascular guidewire in vivo. J Magn Reson Imaging 8:220–225
30. Jolesz FA, Nabavi A, Kikinis R (2001) Integration of interventional MRI with computer-assisted surgery. J Magn Reson Imaging 13:69–77
31. Tsekos NV, Khanicheh A, Christoforou E, Mavroidis C (2007) Magnetic resonance-compatible robotic and mechatronics systems for image-guided interventions and rehabilitation: a review study. Annu Rev Biomed Eng 9:351–387
32. Hempel E, Fischer H, Gumb L, Hohn T, Krause H, Voges U, Breitwieser H, Gutmann B, Durke J, Bock M, Melzer A (2003) An MRI-compatible surgical robot for precise radiological interventions. Comput Aided Surg 8:180–191
33. Pfleiderer SO, Reichenbach JR, Azhari T, Marx C, Malich A, Schneider A, Vagner J, Fischer H, Kaiser WA (2003) A manipulator system for 14-gauge large core breast biopsies inside a high-field whole-body MR scanner. J Magn Reson Imaging 17:493–498
34. Koseki Y, Washio T, Chinzei K, Iseki H (2002) Endoscope manipulator for trans-nasal neurosurgery, optimized for and compatible to vertical field open MRI. In: Proc medical image computing and computer-assisted intervention. Springer, Tokyo, Japan, pp 114–121
35. Gronemeyer DH, Seibel RM, Schmidt A, Melzer A, Dell M (1996) Two- and three-dimensional imaging for interventional MRI and CT guidance. Stud Health Technol Inform 29:62–76
36. Scholz M, Deli M, Wildforster U, Wentz K, Recknagel A, Preuschoft H, Harders A (1996) MRI-guided endoscopy in the brain: a feasibility study. Minim Invasive Neurosurg 39:33–37
37. Masamune K, Kobayashi E, Masutani Y, Suzuki M, Dohi T, Iseki H, Takakura K (1995) Development of an MRI-compatible needle insertion manipulator for stereotactic neurosurgery. J Image Guid Surg 1:242–248
38. Kaiser WA, Fischer H, Vagner J, Selig M (2000) Robotic system for biopsy and therapy of breast lesions in a high-field whole-body magnetic resonance tomography unit. Invest Radiol 35:513–519
39. Felden A, Vagner J, Hinz A, Fischer H, Pfleiderer SO, Reichenbach JR, Kaiser WA (2002) ROBITOM-robot for biopsy and therapy of the mamma. Biomed Tech (Berl) 47:2–5
40. Larson BT, Erdman AG, Tsekos NV, Yacoub E, Tsekos PV, Koutlas IG (2004) Design of an MRI-compatible robotic stereotactic device for minimally invasive interventions in the breast. J Biomech Eng 126:458–465
41. Chinzei K, Miller K (2001) Towards MRI guided surgical manipulator. Med Sci Monit 7: 153–163
42. Susil RC, Krieger A, Derbyshire JA, Tanacs A, Whitcomb LL, Fichtinger G, Atalar E (2003) System for MR image-guided prostate interventions: canine study. Radiology 228:886–894
43. Susil RC, Camphausen K, Choyke P, McVeigh ER, Gustafson GS, Ning H, Miller RW, Atalar E, Coleman CN, Menard C (2004) System for prostate brachytherapy and biopsy in a standard 1.5 T MRI scanner. Magn Reson Med 52:683–687
44. Tsekos NV, Ozcan A, Christoforou E (2005) A prototype manipulator for magnetic resonance-guided interventions inside standard cylindrical magnetic resonance imaging scanners. J Biomech Eng 127:972–980
45. Tsekos NV, Christoforou E, Ozcan A (2008) A general-purpose MR-compatible robotic system: implementation and image guidance for performing minimally invasive interventions. IEEE Eng Med Biol Mag 27:51–58

46. Ozcan A, Tsekos NV (2008) The interconnection of MRI scanner and MR compatible robotic device: synergistic graphical user interface to form a mechatronic system. IEEE/ASME Trans Mechatron 13:362–369
47. Christoforou E, Akbudak E, Ozcan A, Karanikolas M, Tsekos NV (2007) Performance of interventions with manipulator-driven real-time MR guidance: implementation and initial in vitro tests. Magn Reson Imaging 25:69–77
48. Karanikolas M, Christoforou E, Akbudak E, Eisebeis PE, Tsekos NV (2006) An archetype for MRI guided tele-interventions. International Federation for Information Processing. pp 476–483
49. Christoforou E, Tsekos NV (2006) Robotic manipulators with remotely-actuated joints: implementation using drive shafts and u-joints. In: Proc IEEE int conf on robotics and automation. Orlando, FL, pp 2886–2871
50. Pfleiderer SO, Marx C, Vagner J, Franke RP, Reichenbach JR, Kaiser WA (2005) Magnetic resonance-guided large-core breast biopsy inside a 1.5-T magnetic resonance scanner using an automatic system: in vitro experiments and preliminary clinical experience in four patients. Invest Radiol 40:458–463
51. McVeigh ER, Guttman MA, Lederman RJ, Li M, Kocaturk O, Hunt T, Kozlov S, Horvath KA (2006) Real-time interactive MRI-guided cardiac surgery: aortic valve replacement using a direct apical approach. Magn Reson Med 56:958–964
52. Hempel E, Fischer H, Gumb L, Hohn T, Krause H, Voges U, Breitwieser H, Gutmann B, Durke J, Bock M, Melzer A (2003) An MRI-compatible surgical robot for precise radiological interventions. Comput Aided Surg 8:180–191
53. Karanikolas M, Christoforou E, Akbudak E, Eisebeis PE, Tsekos NV (2006) An archetype for MRI guided tele-interventions. In: International Federation for Information Processing pp 476–483
54. Akbudak E, Zuehlsdorff S, Christoforou E, Ozcan A, Karanikolas M, Tsekos NV (2006) Free-hand performance of interventions with manipulator-driven real-time dynamic update of the imaging plane. In: Proceedings of the int soc of magn reson med 14th meeting and exhibition. Seattle, p 1443
55. Jolesz FA (2005) Future perspectives for intraoperative MRI. Neurosurg Clin N Am 16: 201–213

Image-Guided Interventions and Robotics

Bernard Bayle, Oliver Piccin, Laurent Barbé, Pierre Renaud, and Michel de Mathelin

Robot-Assisted Procedures in Interventional Radiology

Interventional Radiology

Procedures

Interventional radiology (IR) is a branch of medicine in which radiologists perform diagnoses or treatments with needles, catheters, or probes, guided by one or several imaging modalities [1]. In many cases, these procedures offer a minimally invasive alternative to traditional surgery. The long list of indications for IR procedures justifies the interest it has aroused in recent years [2].

Vascular interventions are performed with catheters directly introduced by endovenous or endoarterial approach. There are two main categories of treatments: embolization (occlusion) or angioplasty (opening). Embolizations can be performed to stop bleeding, to inject therapeutic drugs, or to provoke tumor ischemia, for example. In angioplasty, narrowed or obstructed vessels are released by inflating balloons located at the tip of a catheter.

Nonvascular interventions are fast-developing procedures. They are performed with needles or needle-shaped tools. They can be used for diagnosis, as the biopsies that consist in the removal of tissue samples for analysis purpose. They can also be used for therapy by guiding tools directly into the target, for instance, to perform tumor ablation. Radiofrequency ablation is an example of such a procedure, performed with a needle-like probe placed into the tumor and used to cook cancer cells. IR also improves the repairing process of fractured vertebra. Using a percutaneous guide inserted in the vertebra under radiography control, bone cement is injected to consolidate the vertebra, which offers an alternative to open orthopedic surgery with less trauma.

B. Bayle (✉), O. Piccin, L. Barbé, P. Renaud, and M. de Mathelin
LSIIT, University of Strasbourg, France
e-mail: bernard.bayle@unistra.fr

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_11,

Imaging

Radiography uses X-ray in diagnosis and medical treatments. Small doses of ionizing radiations are used to produce planar pictures of the inside of the body. Fluoroscopy consists in the continuous acquisition of radiographic images in order to see internal organs in motion. It is performed using a C-shaped device called C-arm that can be rotated around the operation bed. It has to be noticed that robotized fluoroscopes have been developed during the last few years to explore 3D imaging with fluoroscopy. One of them, the Artis Zeego multiaxis system by Siemens [3], which has recently been commercialized, allows to acquire volumes by moving a C-arm around the patient. Associated to volumic image reconstruction, it multiplies the possibilities for diagnoses and interventions.

Ultrasound (US) imaging consists in capturing images of the inside of the patient by local exposure to high-frequency sound waves. It is a nonionizing and real-time modality, which is easy to use and less expensive than other techniques. Additionally, it allows continuous scanning of the organs to observe their physiological motions. It can be used alone in IR, but only for superficial procedures such as liver biopsies. Doppler US imaging can improve the potential use of US as it shows relevant information on the blood flow (blocks, narrowing), but also on tumors [4].

Computed tomography (CT) imaging is based on the acquisition of several two-dimensional X-ray images by a CT scanner. This imaging device associates several beams and detectors, which rotate around the patient while the couch is moving into the scanner gantry. The obtained images characterize the density of the tissues that have been passed through. The very good spatial resolution of the images is CT's most noticeable advantage. It should also be noticed that it can be used for a very wide range of organs. Unfortunately, CT is by far the most ionizing technique.

Magnetic resonance imaging (MRI) uses the magnetic properties of the hydrogen protons to provide a signal, which can be detected and translated into a volumic representation. Though metals are not compatible with the MRI environment because of the high magnetic field, it remains a quite convenient imaging device for IR applications as it is not ionizing. MRI provides very good contrast, even without contrast agents. Unlike other modalities, it also has the particularity to offer direct multiplanar reconstructions.

Table 1 summarizes the pros and cons of these imaging modalities. For each criterion, + indicates an advantage of the imaging modality and − indicates a drawback. Notice that in the field of IR, the different imaging modalities (see Figure 1)

Table 1 Image modalities pros and cons

	US	X	CT	MRI
Imaging quality	−	+	++	++
Variety of targets	−−	−	++	+
Radiations	++	−−	−−	++
Availability	++	+	−	−−
Price	++	−	−−	−−

Fig. 1 IR and multimodal imaging: X, CT, and US (Courtesy MD X. Buy)

are more and more used simultaneously to combine their advantages; for instance, it can be useful to observe a liver tumor with the US, while performing an intervention with a CT scanner. It allows to locate the tumor that is not visible without the injection of contrast agents.

Robotized Interventional Radiology

Why Robots for Interventional Radiology?

Positioning and Registration

The interest of robotized percutaneous procedures lies in the safety and accuracy that can be provided: the coupling of the imaging device with the robot can help the radiologist to position accurately the surgical tool, with an increased safety. In fact, this statement is still controversial as a robotic system is not necessarily superior to a human expert. Depending on the complexity of the procedure, radiologists may obtain better results than most mechanical systems whose accuracy depends on mechanical backlash and deformations. However, in some cases, it might be very difficult for the radiologist to estimate the needle travel throughout the images and so the resulting manual accuracy is likely to be poor.

In fact, robotization offers some other less trivial, but interesting advantages. It is the case of automatic registration, which consists in the computation of the transformation between the robot frame and the image frame. Associated to a robotic device, it allows to position automatically the system to the necessary pose so as to reach targets observed in the images.

CT-Guided Procedures

The main drawback of X-ray imaging in IR is the amount of radiations imposed to the physicians when they perform interventions. It leads to wear uncomfortable lead protections. It also naturally limits the number of interventions, of images of each patient, and as a consequence the development of these promising techniques. Manual IR is then rather incompatible with real-time imaging, whereas fluoroscopic CT might be desirable in some cases. As a result, radiologists use images that are acquired from time to time and they perform sequential procedures, increasing the number of acquisitions if necessary. The perception of the forces applied to the instrument thus becomes fundamental for the physician; it helps to understand the nature of the tissues and the transitions between the layers.

The main contribution of robotics is then quite obvious for CT-guided procedures. The remote control of a robot equipped with a needle driver makes it possible to place the radiologist away from the radiations of the imaging device. In this context, most robotic devices proposed for CT imaging have been designed to be teleoperated [5,6]. Two options are possible. If imaging is real-time the remote insertion could possibly be performed without force feedback. But if the imaging protocol is not modified, then force feedback has to be rendered to provide the physician with the same tactile clues it usually has.

MRI-Guided Procedures

MRI is superior to other imaging modalities in terms of soft tissue contrast. It has therefore a great interest for the determination of pathologic tissues, and as a consequence for percutaneous procedures such as biopsies or tumor ablation. MRI-guided procedures are all the more interesting than MRI lacks ionizing radiations. The radiologist can thus remain close to the patient and the scanner without any harmful effects. The interest of teleoperated systems and more generally of robotic devices thus does not lie in the improved safety of the radiologist.

The robotic approach initially appeared necessary in the development of MRI-guided procedures because of accessibility constraints. The highest image level of details is obtained with close-bore scanners for which the patient is placed in a tunnel during the imaging. Until recently, the size of this bore strictly limited the manual procedures, since its diameter was in the order of 60 cm [7] and its length almost comparable to the patient body length. As a consequence, teleoperated systems have been developed to allow the introduction of surgical tools in this confined space

[8,9]. The problem is now less sensitive with the development of open-bore scanners that combine a high magnetic field, in the order of 1.5 T, with a bore diameter above 70 cm and a length of 120 cm. Manual procedures can now be achieved as clinical routines on the kidney, liver, torso, or head [10]. Robotics can still interestingly make MRI-guided procedures more feasible and practical. Indeed, the scanner is still much longer than a CT scanner, which complicates the access to the patient.

Technological Issues

Safety

Robotic devices for IR fall into two major categories, namely, table-mounted and body-mounted systems. A significant safety issue is the risk of tissue laceration due to physiological motions of the organs or patient motions. This risk is obvious in the case of table-mounted robotic systems. The combined use of embedded force sensors and a quick release mechanism for the needle may reduce the risks of injury [11], but a patient-mounted system seems inherently safer since it provides a partial compensation for respiratory and gross motions of the patient. However, it does not completely prevent from tissue damage. Indeed, patient-mounted systems move with the body surface on which they are attached. So they are efficient to prevent the effects of accidental coughing for instance. Conversely, the direction of the organs motions because of respiration is not necessarily the same as the direction of the needle tip point. For instance, during respiration, the liver is pushed by the diaphragm. As a result, the organ and the abdomen surface do not move in the same directions.

Robotic systems for CT: radiation-free procedures for the physician?

As previously underlined, the main originality of robotic assistance in the CT scanner gantry lies in the ability to perform teleoperated needle insertions. To that purpose, two main problems have to be solved: needle driving and force sensing.

Materials. The compatibility of materials with X-ray is mainly related to image artifacts [12]. When metallic materials are placed in the CT scanner, streak artifacts appear in the images because metals density is not compatible with CT. It provokes an important deterioration generally incompatible with the normal interpretation of the images. Then, the use of plastic materials or carbon has to be preferred. It is important to notice that metallic parts can however be used to build the robot, as long as they do not cross the imaging plane. Conventional electromagnetic or piezoelectric actuators can therefore be integrated if they are placed away from the imaging plane [13].

Force feedback. Force feedback is an important issue with CT compatible robots. It becomes critical in the design of systems dedicated to insert needles. Measuring insertion forces imposes to comply with some requirements, including the limited available space and the use of standard needles of varying sizes. In the first generation of needle drivers [14], the working principle of needle insertion

devices was based on friction so that the resulting traction on the needle shaft highly depended on friction conditions, which are not precisely controllable. Then the insertion force could only be measured with a poor accuracy. Moreover, the needle rotation about its axis could not be imposed, whereas it is an interesting feature to perform bevel tip steering and compensate for needle bending [15].

More recently, two systems adapted to force measurement have appeared. The first one is the CT-Bot needle driver [6], based on successive grasping and release of the needle with chucks attached to a translating carriage (it is detailed in Sect. "CT-Bot: A Body-Mounted System for Interventional Radiology Under CT-Guidance"). An alternative system [11], called revolving needle driver (RND), has been also proposed. It uses the quasi linear motion of a rotating arm over a limited range of rotation. Both systems are designed to allow force sensing, which becomes possible when the needle is tightened during the insertion.

Robotic Systems for MRI: A Technological Challenge

Because of the strong magnetic field, all the design basics have to be reconsidered in the case of MRI: sensors, actuators selection as well as the choice of structural materials have been investigated and solutions have been often redeveloped for that context.

Materials. The compatibility of materials with the MR scanner has been quite extensively studied [16]. Some particular metallic alloys can be considered, such beryllium–copper alloys. For the best device transparency, plastic materials (Glass fiber-reinforced polyester [8], Delrin [17], Peek [18]) constitute probably the best choices. In terms of stiffness, they are however by far less interesting than any aluminum or steel alloy. As a consequence, robotic devices developed up to now tend to be more bulky than devices developed in a CT context.

Actuators. Conventional electromagnetic actuators, even with shielding, can only be integrated if they are located far from the scanner, i.e., several meters away. Most preferably, ultrasonic motors can be used, since nonmagnetic versions are commercially available. They, however, affect the image quality. Other technologies have thus been revisited to propose alternative actuation methods. Pneumatic technology lacks compactness, but offers a high compatibility with MRI. Simple pneumatic cylinders [19], or specifically developed devices [17, 20] remain of quite large size. Hydraulic actuation, with for instance water as a fluid [21], allows a higher power/weight ratio. Hydraulically driven devices, even of small size, have thus been designed [22]. Because of the limited bandwidth and significant complexity of these two latter technologies, design of actuators for MRI is still an open field, with alternate approaches such as electroactive polymers [23] or dielectric elastomer actuators [24].

Sensors. The robotic system needs to be registrated with respect to the scanner. To do so, fiducial markers based on gadolinium [25] can be integrated to the structure. The image of the markers is then necessary to perform the computation of the

robot pose with respect to the scanner. An alternate approach is to use the gradient fields of the scanner to determine the position and orientation of miniature coils [26]. In the same time, proprioceptive information is needed to control the robot device, i.e., determine the tool position with respect to the base. Position sensing can be performed with the optical principle used for conventional incremental encoders. The electronics need however to be placed outside the scanner room. Force and torque measurements can also be required if a teleoperated system with haptic feedback is considered. Unfortunately, conventional force/torque sensors are intrinsically not compatible with MRI. Novel components have therefore been developed, mainly using optical sensors [18, 27].

Examples of Image-Guided Robotic Systems

Many robotic systems dedicated to IR have been developed during the last 10 years. In this section, we present a selection of systems that give a good overview of the State of the Art.

The AcuBot system has been developed in the URobotics Laboratory at the Johns Hopkins University (JHU), Baltimore, USA [5], to perform teleoperated percutaneous interventions under CT guidance, particularly for access to the kidney. It is based on the RCM (remote center of motion)-PAKY system. The needle is loaded in the PAKY tool [14], a plastic X-ray compatible needle driver placed at the tip of the RCM module, which is used to define the needle orientation [28]. A 7-degree-of-freedom (DOF) passive arm is used to perform the positioning of the RCM module. It is in turn mounted onto a 3-DOF Cartesian stage for the accurate positioning in XYZ positions. Needle insertion can be performed thanks to a joystick (without force feedback) or automatically, under CT fluoroscopy monitoring [5]. Several clinical studies on cadavers and patients have shown the efficiency of the system. Recently, Shah et al. [11] have presented the RND needle driver (see Sect. "Technological Issues"), which can be connected to the AcuBot. The insertion of the needle and its spinning are fully actuated. For safety reasons, the needle can be released at any time and a force sensor has been integrated into its structure to render force feedback. This new system has been tested in a multimodality imaging clinical environment on a swine model.

The MrBot has been developed by the same laboratory of JHU (see Figure 2a). It is dedicated to prostate interventions under MRI guidance [29, 30], but can be used with any other medical imaging device [31]. It is built exclusively from amagnetic materials, which required the design of new pneumatic stepper motors [20], whose positions are measured by optical sensors. The end-effector of the MrBot can be adapted to various percutaneous instruments.

The German company Innomedic (now Synthes) issued the only ever commercialized needle placement system, called the Innomotion [32]. The system is based on a robotic arm actuated by pneumatic motors and attached to a gantry crane mounted onto the operation table. It can be passively prepositioned according to the region of interest. The measurement of the position is achieved by optical sensors compatible with both CT and MR. The needle trajectory is completely defined from

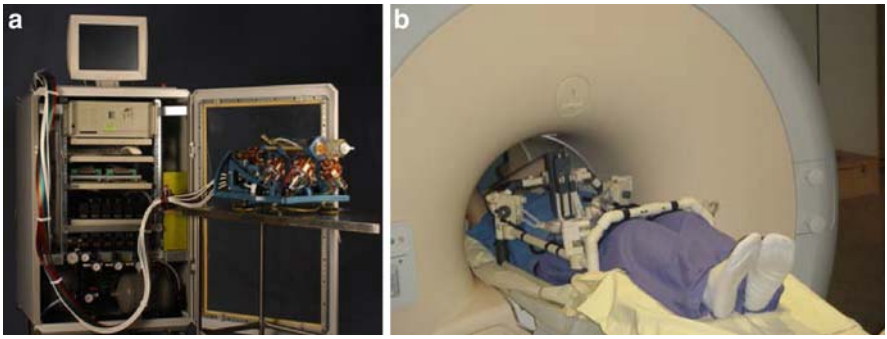


Fig. 2 (**a**) MrBot (courtesy Dr. D. Stoianovici), (**b**) LPR (Courtesy Pr. P. Cinquin)

the pre- and peroperative images by two points: one entry point on the skin surface and one target point inside the body of the patient. The practician selects these points through a graphical user interface and the robot moves automatically toward the referenced trajectory. Finally, the needle is manually inserted. Innomotion has been successfully included in clinical protocol and the performances of the system have been verified by some ex vivo and in vivo experiment. In ref. 8, the authors present a prototype of teleoperated needle insertion based on a modified Innomotion system, but this device is not commercialized yet.

The Light Puncture Robot (LPR, see Figure 2b) has been designed by the TIMC laboratory, Grenoble, France [33], for MRI-guided percutaneous interventions. It is made of plastic materials and actuated by pneumatic motors that ensure MRI compatibility and also limit artifacts when used with CT imaging. The LPR is a 5-DOF robot: four DOF for the positioning and the orientation of the needle and one for the insertion. The robot is patient-mounted and positioned on the abdomen by motorized straps. The needle insertion is performed with the help of a fast linear pneumatic actuator, which is used to perform a single fast insertion motion, in order to perforate the skin and limit the organs motions [17].

The CT-Bot system (see Figure 3b) is a robot dedicated to teleoperated CT-guided percutaneous interventions with force feedback, developed by the LSIIT laboratory, University of Strasbourg, France [13]. It is patient-mounted, such as the LPR, with a special effort on needle driving capabilities [6]. This system is the subject of the next section.

The main characteristics of these systems are mentioned in Table 2.

CT-Bot: A Body-Mounted System for Interventional Radiology Under CT-Guidance

The CT-Bot comprises two main parts: a robotic positioning assistant and a needle insertion system. This architecture corresponds to a decomposition of the needle insertion into two steps, to maximize safety. In the first step, the initial positioning

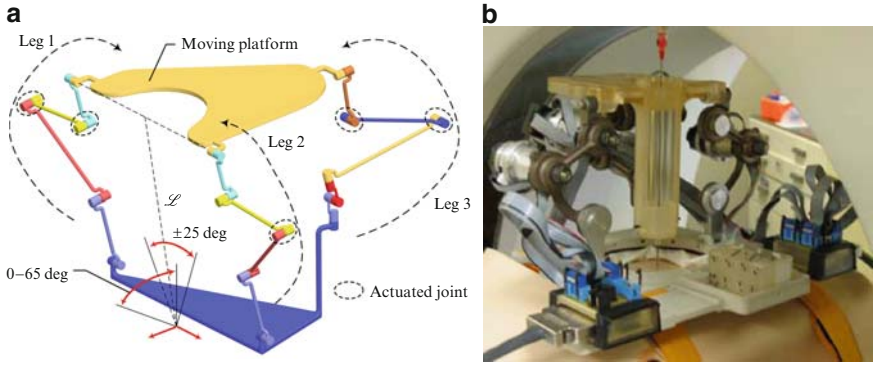


Fig. 3 (**a**) Schematic view of the CT-Bot, (**b**) Prototype of the mechanism in the CT-scanner ring

Table 2 Robotic systems for IR [6]

System	Location	Images	Motion	Validation	Type	Insertion
AcuBot	JHU	CT	IG, TEL	IV	TM	PAKY: friction RND: motorized
MrBot	JHU	CT, MRI	IV	Ph	TM	Pneumatic
Innomotion	Innomedic	CT, MRI	IG	IV	TM	Manual
LPR	TIMC	CT, MRI	IG	IV	PM	Pneumatic
CT-Bot	LSIIT	CT	IG, TEL	Ph: manual IV: robotized	PM	Manual motorized

IG image guidance, TEL teleoperation, Ph phantom validation, IV in vivo validation, TM table-mounted, PM patient-mounted

and orientation of the needle in the free space is performed toward the target. In the second step, the needle insertion is achieved. In this section, we present the main features of the two modules.

Positioning Module

During an intervention, the patient is placed within the gantry of the CT scanner. A large needle incidence is needed to reach the organs: typically 0–65° in the transverse plane and ±25° in the sagittal plane. Another important feature is the possibility to change the entry point position on the patient's skin within a range of at least ±10 mm around a predefined point. Additionally, the positioning should withstand an insertion force in the order of 20 N exerted on the needle axis during the insertion task.

As five DOF are required to position and orient the needle axis in space, we considered existing mechanical architectures with three translational and two rotational mobilities. Both classical serial and parallel mechanisms with fully symmetrical architectures were found to be inappropriate either because of their inherent

flexibility or because of a too limited workspace. Parallel mechanisms with lower mobility were possible candidates for the application, but the specific requirements in terms of motion ranges led us to consider asymmetric mechanisms with different types of legs. From a kinematic viewpoint, the most demanding specifications for the mechanism are the specified needle rotations. The final mechanical architecture selection was synthesized to obtain a three-legged parallel mechanism with five DOF. The first two legs are connected to the moving platform to obtain a planar linkage. The third leg, composed by a spatial mechanism, can swivel simultaneously the moving platform and the planar linkage. The schematic view of the system in Figure 3a shows the CT-Bot architecture with the line $\mathcal{L}$ representing the needle. The asymmetric leg arrangement provides a large orientation capability as well as an easy access to the needle anytime during the insertion procedure.

Figure 3b shows the fabricated prototype of the mechanism in the CT scanner ring with a passive needle guide mounted on the platform during an experimental evaluation of the system [13].

Needle Insertion Module

The needle driver is mounted on the positioning module, and is also compatible with the specifications imposed by IR and CT scanner use. It addresses the insertion force measurement issue and features a novel driving principle based on successive grasping and translations of the needle.

Design Requirements and Working Principles

It is supposed that the entry point and the insertion axis are known and supplied the positioning module to which the needle driver is connected. The main functional characteristics for a robotized percutaneous therapy scenario are summarized in Figure 4, which presents the working principles used to fulfill the requirements.

The needle driver must fit within the bore of the CT scanner, between the patient and the top of the scanner ring. To perform the whole needle translation in the limited available volume, the translational movement is split into a sum of elementary translations. These elementary displacements are obtained using translations of a carriage holding a chuck-like grasping device. Before starting, a small skin incision is made on the patient's skin at the entry point to facilitate the needle insertion task. The needle is then manipulated by a single chuck using the entry point. During the insertion cycle, the chuck is translated back and forth while the needle is grasped and released. The insertion cycle is carried out until the required insertion length is reached. The needle extraction is similar, except that the carriage stands initially in its lower position. Notice that the presented insertion scheme allows the free motions of the needle about the entry point, provided that the chuck yields a wide opening.

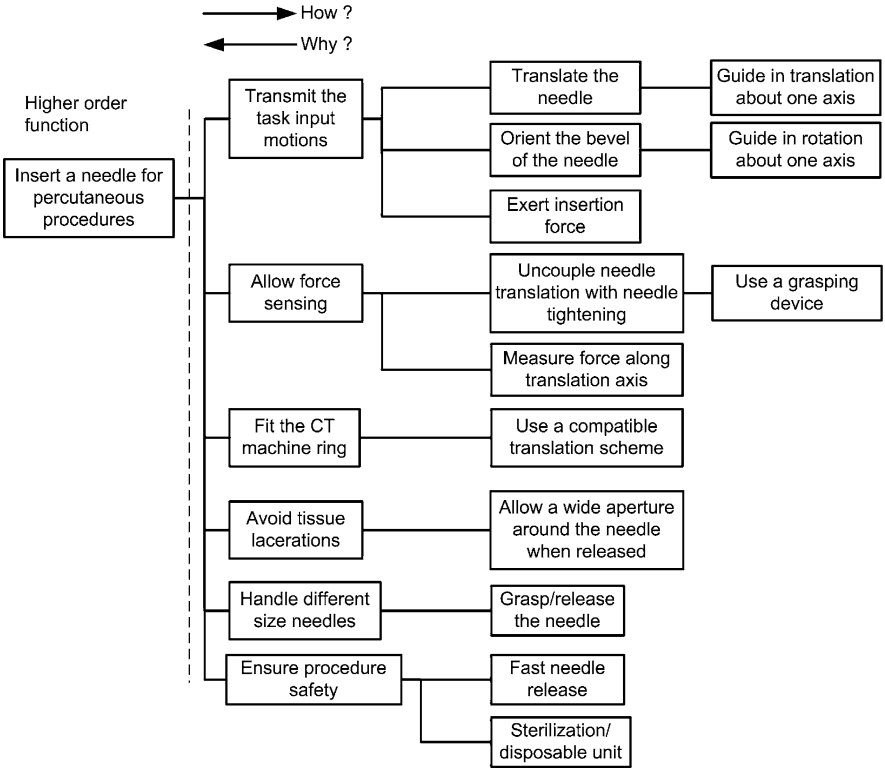


Fig. 4 Functional diagram of the needle insertion module

Safety Issues

As for any medical application, safety has been a top-rank priority concern during the design process. The device force capability depends on the maximum axial effort required. It may not exceed 20 N, as it is limited by the friction occurring between the needle barrel and the jaws, thus providing an inherently safe behavior. In the same line, the decoupling of the insertion from the needle positioning limits the number of active DOF during the most critical phase when the needle is inserted. The insertion task is performed with only one degree of freedom, which reduces significantly the complexity of the safety issue. As a patient-mounted robotic device, the CT-Bot and its insertion tool also provide a safe and simple solution to compensate partly for external patient's motions. Another prominent concern is the effect of respiratory motions on internal organs, which can cause the inserted needle to wobble about the entry point. To avoid tissue laceration, the specifically designed chuck offers a 16 mm opening, which allows the needle to move freely approximately 20° about its central position.

Prototype and In Vivo Experimental Results

A prototype of the needle insertion module has been constructed to carry out teleoperated needle insertions with force feedback. The proposed embodiment addresses the medical requirement of sterilization. The active parts of the system can be wrapped in sterile bags, whereas the passive components can be sterilized and used as disposable units. The experimental setup comprises the slave and the master manipulator visible in Figure 5. The slave manipulator is composed of the needle insertion device mounted on a passive parallel structure, which is attached to an anesthetized swine. The master manipulator is a linear haptic device designed to meet the prescribed constraints of IR [34]. In particular, it has a needle-like grasp and thus can be seized as a conventional needle. The needles used during experiments are classical 18-gauge, 152.4-mm-long biopsy needles.

The experiments consisted in performing a teleoperated needle insertion into the liver of the animal, through an incision on the skin. To reach the desired depth of approximately 100 mm, two insertion cycles are required. Figure 6 represents force and position tracking between the master and the slave. At time $t = 3.1$ s, the needle is grasped and the insertion begins. During the needle insertion, tissue layers are punctured which is evidenced by discontinuities in the force plot at $t = 12.7$, 19.5, and 33.0 s. When the carriage of the needle driver reaches its lower position, the operator hand is constrained by adequate force feedback to remain motionless. In the same time, the carriage moves back to the upper position. Then, a second insertion cycle begins until the operator stops the insertion and releases the needle. The chuck is finally opened and the needle manually extracted. During the whole insertion, the RMS force tracking error remains below 0.5 N. During the transient period between the two insertion phases, when the virtual force feedback is applied to the master, no force is measured at the slave side. This explains the difference between F_h and F_e. On the other hand, position tracking error increases rapidly

Fig. 5 Teleoperated needle insertion experiment

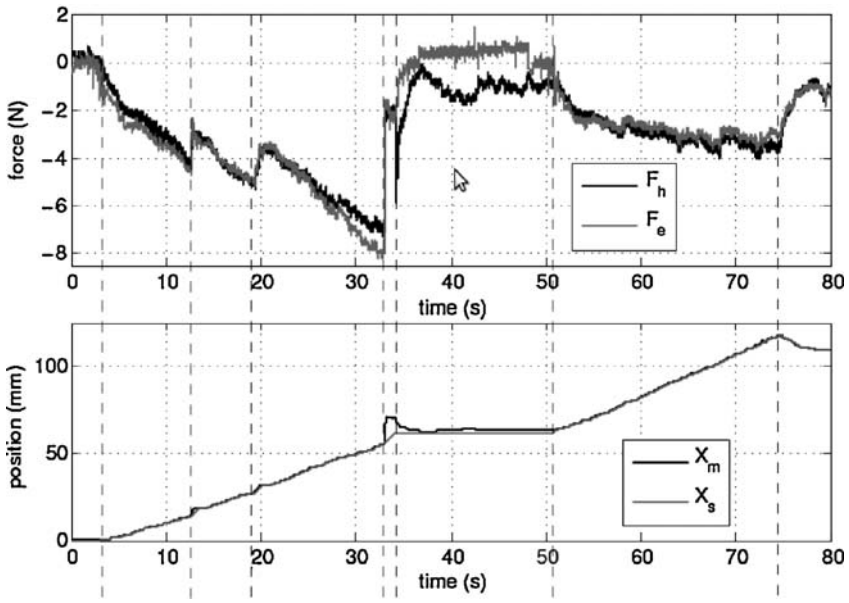


Fig. 6 Needle insertion experiment: force and position tracking [6]

when the needle penetrates the tissues abruptly, but with less than 2 mm during the two insertion phases. Solutions can be proposed to tackle this drawback by adequate detection of such events [35].

Discussion

This brief overview of the challenges of robotic systems for IR leads to several conclusions.

CT Robotics

Robots dedicated to CT scanner have appeared at the end of the 1990s [14]. They led to scientific and technological developments, in particular in the field of registration and design. They raised the problem of safety associated to needle insertions which resulted in two main approaches in the community: table-mounted and body-mounted systems. Though this latter class of robots aims to limit the effects of tissues motions, the problem remains relatively open. Until new innovations, we can consider that active compensation of physiological motions in association with these systems has to be explored more extensively. Force feedback is the other issue that requires further research, as pioneer systems [6, 30] have just emerged after years of developments.

MRI Robotics

Many developments have been necessary to consider the use of robots in association with MRI scanners. Several systems have now been tested and one has been commercialized [8]. From a technological point of view, the next challenge is probably to progress toward lighter, less cumbersome systems for an easier integration in the MR room. From a medical point of view, the needed degree of assistance for MRI-guided percutaneous procedures is still an open question. For certain gestures, switching from a teleoperated system, such as [8] to a collaborative system such as [36], might be relevant in order to combine the assistance in the positioning to the natural haptic feedback that the radiologist can get if he performs manually the insertion gesture. Medical evaluation is still needed to make a decision on that point.

References

1. Kaufman JA, Lee MJ (2004) Vascular and interventional radiology – The requesites. Mosby
2. Society of interventional radiology (2008) http://www.scvir.org
3. Zeego. (2008) The artis zeego multi-axis system. http://www.medical.siemens. com
4. ACR-RNSA. (2008) Radiology info. http://www.radiologyinfo.org
5. Stoianovici D, Cleary K, Patriciu A, Mazilu D, Stanimir A, Craciunoiu N, Watson V, Kavoussi L (2003) Acubot: a robot for radiological interventions. IEEE Trans Robot Automat 19:927–930
6. Piccin O, Barbé L, Bayle B, de Mathelin M, Gangi A (2009) Force feedback teleoperated needle insertion device for percutaneous procedures. Int J Rob Res 28(9):1154–1168
7. Tsekos NV, Chrisotoforou E, Ozcan A (2008) A general-purpose MR-compatible robotic system. IEEE Eng Med Biol Mag 27:51–58
8. Melzer A, Gutmann B, Remmle T, Wolf R, Luboscheck A, Block M, Barden-heuer H, Fischer H (2008) INNOMOTION for percutaneous image-guided interventions. IEEE Eng Med Biol Mag 27:66–73
9. Zemiti N, Bricault I, Fouard C, Sanchez B, Cinquin P (2008) LPR: A CT and MR-compatible puncture robot to enhance accuracy and safety of image-guided interventions. IEEE/ASME Trans Mechatron 13:306–315
10. Ahrar J, Javadi S, Valenzuela Y, Gupta S, Stafford R, Ahrar K (2008) MRI-guided biopsy using a high field strength magnet: Initial experience with 95 patients. In 94th Scientific Assembly and Annual Meeting of the Radiological Society of North America, Chicago, USA
11. Shah S, Kapoor A, Ding J, Guion P, Petrisor D, Karanian J, Pritchard WF, Stoianovici D, Wood B, Cleary K (2008) Robotically assisted needle driver: evaluation of safety release, force profiles, and needle spin in a swine abdominal model. Int J Comput Assist Radiol Surg 3: 173–179
12. Barrett J, Keat N (2003) Artefacts in ct: recognition and avoidance. ImPACT, London, UK. http://www.impactscan.org/slides/rsna2003/ctartefacts.pdf
13. Maurin B, Bayle B, Piccin O, Gangloff J, de Mathelin M, Doignon C, Zanne P, Gangi A (2008) A patient-mounted robotic platform for CT-scan guided procedures. IEEE Transact Biomed Eng 55:2417–2425
14. Stoianovici D, Cadeddu J, Demaree R, Basile H, Taylor R, Whitcomb L, Kavoussi L (1997) A novel mechanical transmission applied to percutaneous renal access. Proceedings of the ASME Dynamic Systems and Control Division 61:401–406

15. Webster RJ, Kim JS, Cowan NJ, Chirikjian GS, Okamura AM (2006) Nonholo-nomic modeling of needle steering. Int J Robot Res 25:509–525
16. Chinzei K, Kikinis R, Jolesz F (1999) MR compatibility of mechatronic devices: Design criteria. Med Image Comput Comput Assist Interv 1679:1020–1031
17. Bricault I, Jauniaux E, Zemiti N, Fouard C, Taillant E, Dorandeu F, Cinquin P (2008) LPR: a light puncture robot for CT and MRI interventions. IEEE Eng Med Biol Magzine 27:42–50
18. Tada M, Kanade T (2005) Design of an MR-compatible three-axis force sensor. IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada, 2618–2623
19. Fischer G, Iordachita I, Csoma C, Tokuda J, DiMaio S, Tempany C, Hata N, Fichtinger G (2008) MRI-compatible pneumatic robot for transperineal prostate needle placement. IEEE/ASME Trans Mechatron 13:295–305
20. Stoianovici D, Patriciu A, Petrisor D, Mazilu D, Kavoussi L (2007) A new type of motor: pneumatic step motor. IEEE/ASME Trans Mechatron 12:98–106
21. Ganesh G, Gassert R, Burdet E, Bleuler H (2004) Dynamics and control of an MRI compatible master-slave system with hydrostatic transmission. IEEE International Conference on Robotics and Automation, New-Orleans, USA, 1288–1294
22. Okayasu H, Okamoto J, Fujie M, Umezu M, Iseki H (2003) Development of a hydraulic-driven flexible manipulator for neurosurgery. CARS 1256:607–613
23. Tadakuma K, DeVita L, Plante J, Shaoze Y, Dubowsky S (2005) The experimental study of a precision parallel manipulator with binary actuation: with application to MRI cancer treatment. IEEE International Conference on Robotics and Automation, Pasadena, USA, 2503–2508
24. Carpi F, Khanicheh A, Mavroidis C, DeRossi D (2008) MRI compatibilityof silicone-made contractile dielectric elastomer actuators. IEEE/ASME Trans Mechatron 13:370–374
25. Beekley (2009) http://www.beekley.com
26. Robin medical (2009) http://www.robinmedical.com
27. Chapuis D, Gassert R, Sache L, Burdet E, Bleuler H (2004) Design of a simple MRI/fMRI compatible force/torque sensor. IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, 2593–2599
28. Stoianovici D, Whitcomb L, Anderson J, Taylor R, Kavoussi L (1998) A modular surgical robotic system for image guided percutaneous procedures. Medical Image Computing and Computer-Assisted Intervention, Cambridge, USA
29. Muntener M, Patriciu A, Petrisor D, Mazilu D, Bagga H, Kavoussi L, Cleary K, Stoianovici D (2006) Magnetic resonance imaging compatible robotic system for fully automated brachytherapy seed placement. Urology 68:1313–1317
30. Muntener M, Patriciu A, Petrisor D, Schar M, Ursu D, Song D, Stoianovici D (2008) Transperineal prostate intervention: robot for fully automated MR imaging–system description and proof of principle in a canine model. Radiology 247:543–549
31. Cleary K, Melzer A, Watson V, Kronreif G, Stoianovici D (2006) Interventional robotic systems: applications and technology state-of-the-art. Minim Invasive Ther Allied Technol 15:101–113
32. Hempel E, Fischer H, Gumb L, Hohn T, Krause H, Voges U, Breitwieser H, Gutmann B, Durke J, Bock M, Melzer A (2003) An MRI-compatible surgical robot for precise radiological interventions. Comput Aided Surg 8:180–191
33. Taillant E, Avila-Vilchis J-C, Bricault I, Cinquin P (2004) CT and MR compatible light puncture robot: architectural design and first experiments. Medical Image Computing and Computer-Assisted Intervention, Saint-Malo, France, 145–152
34. Barbé L, Bayle B, Piccin O, Gangloff J, de Mathelin M (2007) Design and evaluation of a linear haptic device. In IEEE Conference on Robotics and Automation, Roma, Italy
35. Barbé L, Bayle B, de Mathelin M, Gangi A (2007) In vivo model estimation and haptic characterization of needle insertions. Int J Robot Res 26:1283–1301
36. Kronreif G, Fürst M Ptacek W, Kornfeld M, Kettenbach J (2006) Robotic system for image guided therapy – B-robii. International Workshop on Robotics in Alpe-Adira-Danube Region, Balatonfüred Lake Balaton, Hungary

Part IV
Modeling, Simulation and Experimental Data

Emerging Mechanisms of Vein Graft Failure: The Dynamic Interaction of Hemodynamics and the Vascular Response to Injury

Scott A. Berceli, Roger Tran-Son-Tay, and Marc Garbey

Limited Durability of Vein Bypass Grafts

Recently published data demonstrate that approximately 8 million men and women over 40 years of age have peripheral arterial disease and approximately, 20% of Americans greater than 65 years of age endure the symptoms of this disease process [1]. Despite its prevalence and cardiovascular risk implications, durable treatment options are limited, with only about 25% of PAD patients currently undergoing treatment. [2] Surgical revascularization using autologous vein remains among the dominant therapeutic options, with 110,000 peripheral bypass procedures performed in 2004. [1] Fueled by an epidemic of obesity and diabetes in the United States, substantial increases in the need for these interventions are projected over the next decade. Despite the escalating need for these often limb- and life-saving procedures, their medium and long-term durability remains compromised. Plagued by the problems of aggressive stenosis and luminal narrowing, contemporary data shows that almost 40% of lower extremity vein bypass grafts develop occlusive lesions or fail within a year (Fig. 1) [3]. Since many of the technical advancements for improved patency have been exhausted, the current concept is that the future of enhancing the durability of these reconstructions lies in a better understanding of the biology of the vein graft wall in response to interventions [4].

S.A. Berceli (✉)
University of Florida, Box 100128, Gainesville, FL, 32610, USA
e-mail: bercesa@surgery.ufl.edu

R. Tran-Son-Tay
University of Florida, Box 116250, Gainesville, FL, 32611, USA
e-mail: rtst@ufl.edu

M. Garbey
University of Houston, 501 Phillip Hoffman Hall, Houston, TX 77204, USA
e-mail: garbey@cs.uh.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_12,

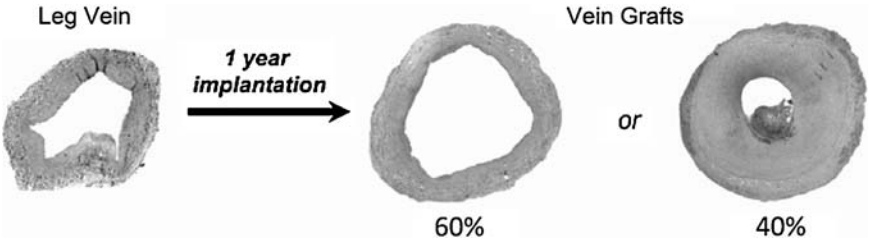


Fig. 1 Vein segments used for arterial bypass grafting demonstrate an adaptive or occlusive phenotype within one year after implantation

Hemodynamics: Modulator of Vascular Adaptation

The hallmark of vascular occlusive disease is the specific localization patterns and focal nature of the disease process. On the basis of morphologic and mechanistic studies, arterial adaptation to altered shear stress encompasses two distinct processes: intimal hyperplasia and wall (inward/outward) remodeling. [5, 6] Intimal hyperplasia is characterized by migration of smooth muscle cells into the intima with proliferation and deposition of extracellular matrix, resulting in narrowing of the arterial lumen. In contrast, remodeling is characterized by preservation or loss of lumen area through reorganization of the cellular and extracellular components within the media. Following catheter-based interventions or surgical manipulation, both forms of adaptation are initiated, and it is the balance between these two processes that dictates the degree of luminal narrowing and ultimately, the success or failure of the intervention. [7] Therapies to improve long-term patency will be achieved through modulation of the mediators that control the balance between these two processes [8].

The acute transposition of a vein segment from a low pressure and flow environment to the high pressure and flow arterial system leads to significant structural changes within the wall (Fig. 2) [9, 10]. These changes are characterized by an increase in both intimal and medial thickness, a burst of smooth muscle cell proliferation with conversion from a contractile to synthetic phenotype, and the extracellular deposition of type I collagen and proteoglycans [11, 12]. These early events in vein graft adaptation frequently continue in an uncontrolled manner, leading to severe lumen narrowing and subsequent graft failure. Vital to understanding these events, however, is the clinical observation that vein graft occlusive disease is segmental, with clear regional variability in the processes that drive the wall towards an adaptive or pathologic phenotype. [13, 14] Analogous to the focal nature of atherosclerotic disease in the arterial system, [15] local wall shearing forces have been postulated as a major regulator of these events.

Research in our laboratory has been focused on using animal models to examine the influence of hemodynamics on vein graft remodeling. [16, 17] Using rabbit carotid vein interposition grafts, coupled with unilateral ligation, we have demonstrated that reductions in blood velocity significantly augment neointimal growth

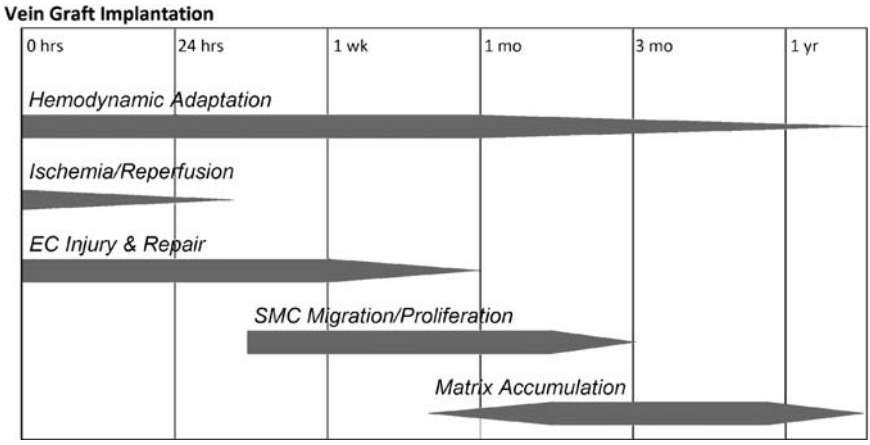


Fig. 2 Overlapping metabolic processes that dictate vein graft remodeling following harvest and implantation. Adapted with permission [4]

(Fig. 3). Complete ligation of the internal carotid artery and partial ligation of the external carotid artery results in an immediate 90% reduction in flow on the ligated (low flow) vein graft side, and an immediate 15-fold difference in wall shear stress between low flow and high flow grafts. Intimal hyperplasia is first noted at day 7, with enhanced intimal thickening in the low flow graft. These differential rates of hyperplasia lead to a sevenfold difference in intimal thickness at day 28. While initially stable and independent of the imposed shear, vein graft diameter (measured at the external elastic lamina) is noted to increase, most prominently under high shear conditions.

Researchers have sought to elucidate the mechanisms for this shear-directed vascular remodeling. Evidence for the direct action of shear stress on endothelial function comes primarily from in vitro studies, where it has been demonstrated that genes important to vascular endothelial physiology, such as nitric oxide synthase (NOS), endothelin-1, fibroblast growth factor (FGF), platelet-derived growth factor (PDGF), and thrombomodulin are all influenced by this stimulus. Several candidate shear stress sensors in endothelial cells have been postulated, which include integrins, ion channels, G proteins, and MAPK coupled receptors. Acting through a shear stress responsive element, consisting of a 6-base pair core element (GAGACC) in many promoter regions, NF-κB, AP-1, Sp-1, and Erg-1 can be activated by changes in shear. Despite these significant insights into the interaction of the endothelium with the flow environment, in vivo evidence of their biologic significance is lacking. Work in our laboratory has been focused on identifying the key signaling pathways that control vascular remodeling in intact vasculature. While we have discovered a handful of mediators to be differentially regulated by flow during early vein graft adaptation, (TNF-α, IL1-β, MMP-2), [18–22] inhibition of these mediators has failed to significantly influence this response. [22] Further underscoring this observation are the recent microarray studies performed in our laboratory,

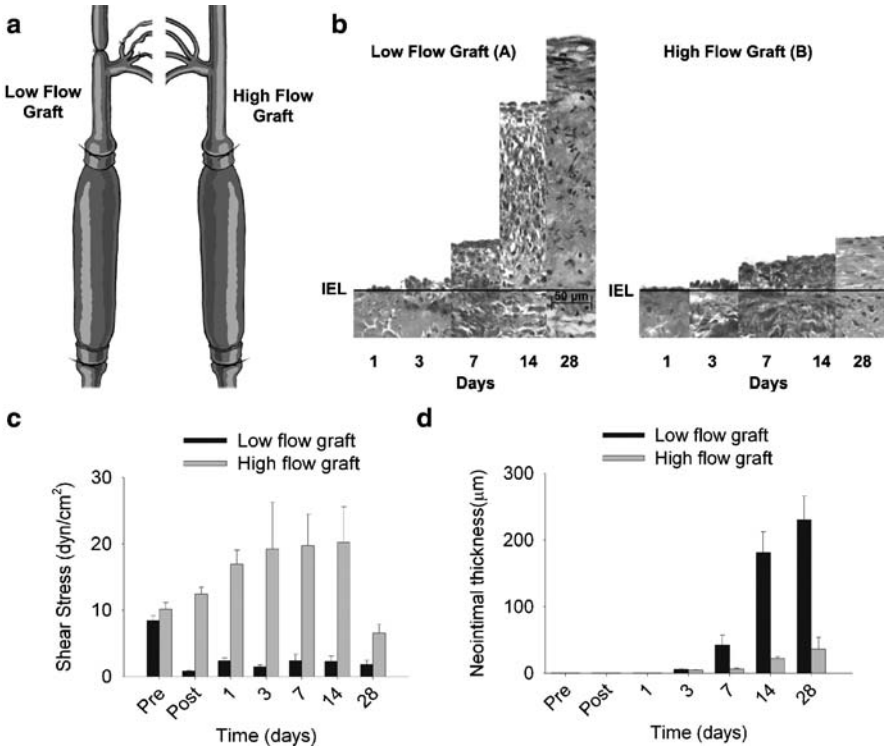


Fig. 3 Effect of flow on intimal hyperplasia development. Using the rabbit bilateral vein graft, partial distal branch ligation construct (**a**), an immediate fivefold reduction in wall shear in the right and twofold increase in wall shear in the left-sided vein graft is induced (**c**). Enhanced intimal hyperplasia development is observed under low flow conditions starting at 7 days (**b**, **d**)

where vein grafts exposed to divergent shear were analyzed for their gene expression patterns. Despite approximately 10% of the genes being up or downregulated during the 28-day implantation period (at a false discovery rate of 0.10), less than 0.1% of these genes were differentially regulated in response to wall shear.

Two central questions regarding occlusive adaptations (inward remodeling and neointimal hyperplasia) following vascular intervention remain unanswered:

- Why does only a subset of patients following vein graft implantation demonstrate pathologic occlusive adaptations leading to lumen narrowing and failure of the graft, while other patients undergoing the same therapy exhibit a favorable vessel wall healing response and preservation of the lumen?
- Why is there substantial regional heterogeneity in vascular repair, such that focal regions of the vein graft wall demonstrate advanced occlusive adaptations while neighboring segments are essentially free of disease?

On the basis of our ongoing investigations, we have hypothesized that the impact of shear stress on vascular adaptation to be more complex than can be described

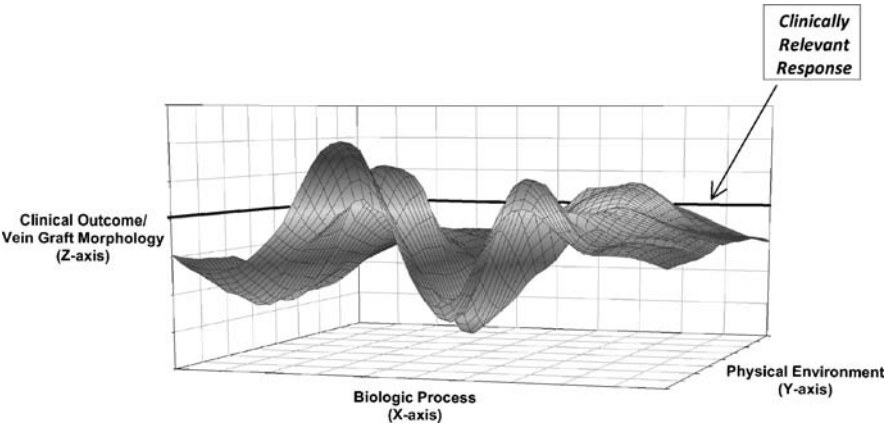


Fig. 4 The remodeling response of vein grafts following implantation is determined by the complex interaction between the biologic process and the physical environment. Specific combinations of these events lead to clinically relevant phenotypes, such as critical narrowing of the graft lumen

through a limited set of key cytokine or signaling mediators. Instead, we theorize that the physical and biologic environments are implicitly linked and lead to specific vein graft phenotypes. Focusing on the critical process of monocyte entry into wall, we have identified that both chemokine upregulation and reduced shear (with increased monocyte surface residence time) are necessary to drive the intimal hyperplastic response. While such simple experiments suggest an important interaction between the physical and biologic microenvironments, a quantitative understanding of how these interconnected networks govern the remodeling process is lacking. It is only through rigorous, analytical evaluation of these various components that we can begin to predict the dynamic response of this complex system to perturbations in local environment (Fig. 4).

Mathematical Modeling of Shear-Mediated Vein Graft Adaptation

In an effort to provide insight into the kinetics of vein graft adaptation, we have developed a mathematical model of vein graft intimal hyperplasia induced by shear stress based on experimental data obtained from the bilateral vein graft construct (Fig. 3). Extending the observations of previous investigators, [12, 23, 24] an inverse correlation between intimal thickness and shear was assumed via the following expression:

$$\frac{\mathrm{d}h}{\mathrm{d}t} = \frac{R_{\text{lumen}} A \mathrm{e}^{-A(t-t^*)}}{1 + B\tau^C} \quad t \geq t^* \tag{1}$$

where h_0 is the intimal thickness at implantation, R_{lumen} is the initial lumen radius, τ is shear stress, t is time, t^* is the time when initial intima is identified, and A, B and, C are empirical constants. Equation (1) was fit to the intimal thickness data points ranging from 1 to 28 days after implantation, and t^* was assigned a value of 3 days, based on the experimental morphology data.

To explore the utility of the predictive model of intimal growth outside the uniform geometry used in its derivation, a rabbit focal vein graft stenosis construct was employed. Due to the complex geometry around the focal stenosis, an analytical solution to estimate shear stress along the length of the graft was not feasible. As such an iterative, numerical approach was used, where sequential calculation of local shear stress, using a finite element approach, and an estimate of the interval intimal growth, computed from equation (1) was performed in a repetitive fashion. Figure 5 shows the path lines and wall shear stress profiles at Day 0 and Day 28 for stenotic grafts exposed to low or high flow conditions. The corresponding experimentally measured and predicted graft geometries ($n = 5$ per group) at Day 28 for each flow simulation. Although the model prediction under low flow conditions demonstrated reasonable accuracy, the model prediction for high flow grafts severely underestimated the extent of disease. The underlying etiology for this accelerated disease process in the high flow grafts is unclear, and this illustrates the dominant weakness of this approach. While we can devise mathematical relationships that correlate vein graft remodeling to the local hemodynamics, they provide us little insight into the biology that dictates these adaptation events, most evident when predictions fall substantially short of experimental data.

System Biology and Vascular Adaptation

While multiscale mathematical models have been extensively applied to the area of cancer research [25, 26] and the inflammatory response to injury, [27, 28] similar applications to the vascular system have been lacking. Estimates of vascular adaptation have been predominately obtained through use of empirical relationships, with no basis in the underlying physiology of the system. [29, 30] Among the few advances in the field has been the work of Serini, who developed and experimentally validated a mathematical model of microvascular growth and reorganization based on the distribution of local chemoattractants as the fundamental mechanism for cell-to-cell communication. [31] More recently, Pierce and her colleagues used an agent based modeling approach to examine the dynamics of leukocyte trafficking in the microvasculature [32, 33].

In a general sense, systems biology encompasses the science of discovery, modeling, understanding, and ultimately, engineering the dynamic relationships that define living organisms [34, 35]. While a wide range of investigative approaches may be encompassed by this definition, Hood and his colleagues at the Institute of Systems Biology have extracted their thoughts of the essential features of contemporary systems biology. [34] Cornerstone to their concepts is the use of quantitative measurement to measure the dynamic changes in response to distinct environmental

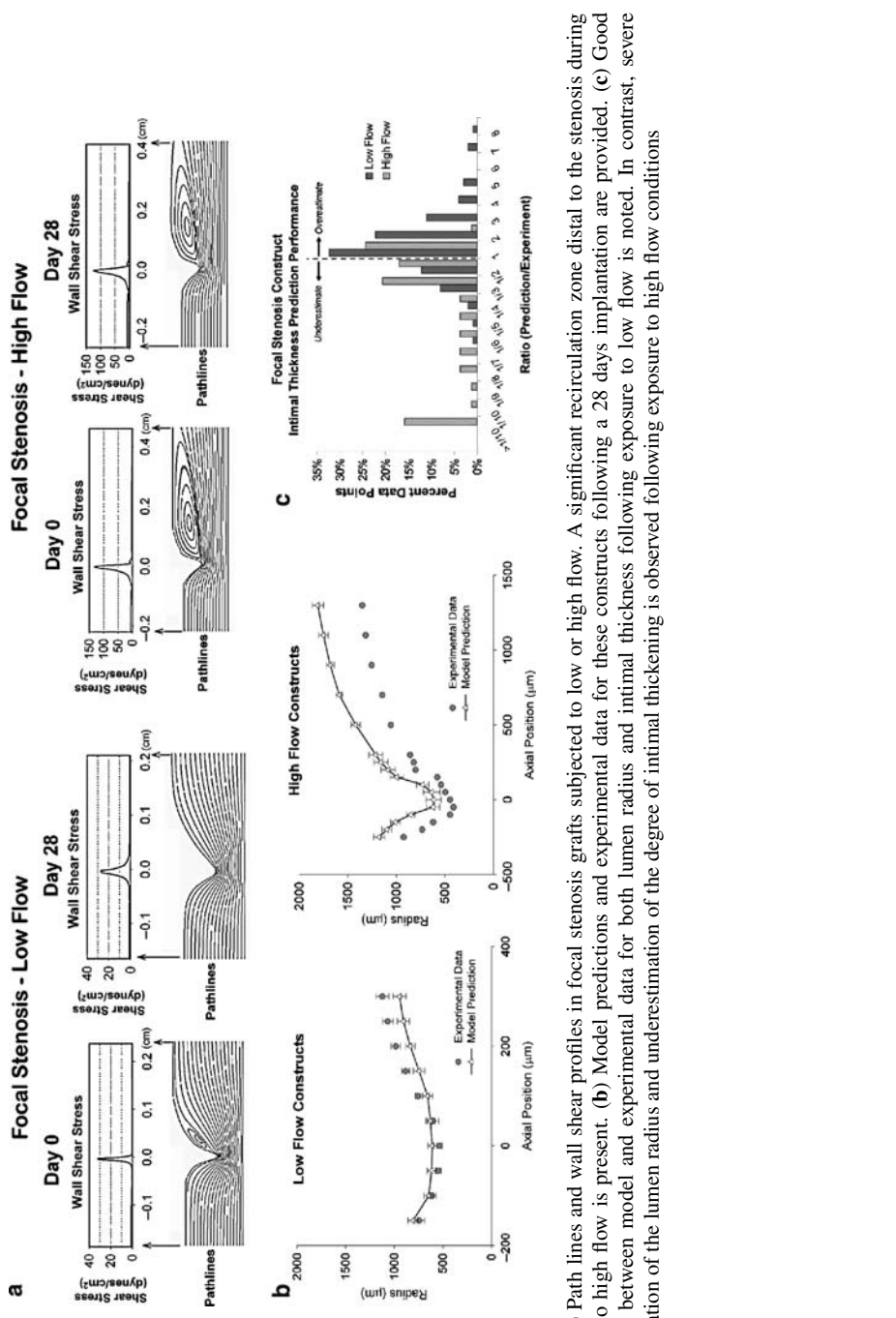


Fig. 5 (**a**) Path lines and wall shear profiles in focal stenosis grafts subjected to low or high flow. A significant recirculation zone distal to the stenosis during exposure to high flow is present. (**b**) Model predictions and experimental data for these constructs following a 28 days implantation are provided. (**c**) Good agreement between model and experimental data for both lumen radius and intimal thickness following exposure to low flow is noted. In contrast, severe overestimation of the lumen radius and underestimation of the degree of intimal thickening is observed following exposure to high flow conditions

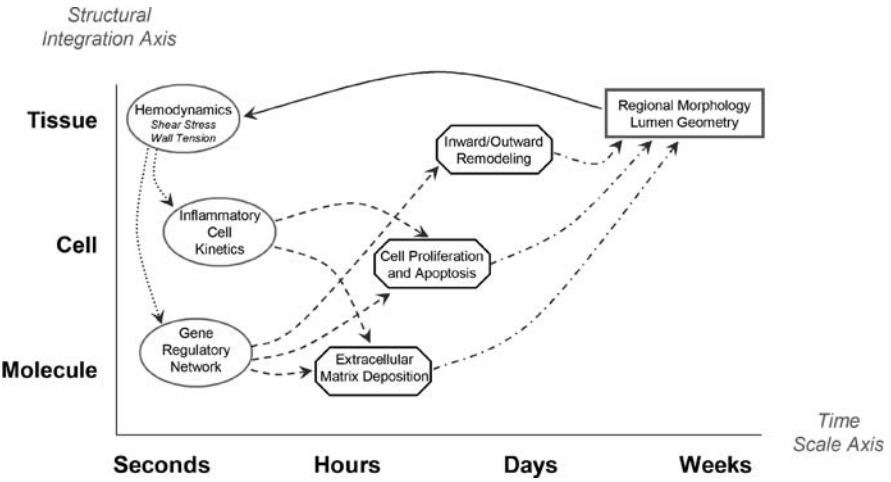


Fig. 6 Multiscale nature of the predictive vascular adaptation model, integrating structural elements from the molecular to tissue level and time from seconds to weeks

perturbations. Our laboratory has initiated the development of a multiscale model of vascular adaptation designed around these concepts (Fig. 6). The dynamic interplay between physical forces, cellular inflammatory elements, and an underlying gene regulatory network is critical to the development of our approach. The resulting model details a highly integrated system where local perturbations in a single component rapidly feedback to the other elements, leading to an updated but stable set point for the network; or a condition with dynamic instability, characterized by early failure of the system. Detailed examination of our model system demonstrates such a critical recursive loop between the local hemodynamics and the regional biologic response of the vascular wall. Initial shear stress not only directs the primary set point for the gene network but also modulates monocyte infiltration, both of which influences the cell and matrix-based remodeling response and defines the local modifications in conduit geometry. These morphologic changes induce perturbations in local shear, resulting in new set points for the biologic response parameters.

Among the principle challenges in the area of cardiovascular research revolves around our ability to translate the extensive body of basic vascular biology into the clinical care of patients. Recent failures in the Corgentech/Bristol-Myers-Squibb development of edifoligide, an E2F transcription factor decoy thought to have a role in the reduction of restenosis following vascular intervention, stand as a glaring example of these challenges. On the basis of successful preclinical studies in an animal vein graft model, human Phase I and II clinical studies were initiated. [36–38] Demonstrating safety with some indication of possible efficacy, [39, 40] two phase III clinical trials were undertaken to investigate the use of edifoligide in improving the durability of vein grafts used in coronary artery or infrainguinal revascularization. Following randomization of 3300 patients and trial expenditures in excess of $40 million, treatment with edifoligide was demonstrated to have no influence on the resulting vein graft morphology.

Analysis of the decision points leading to this costly miscalculation reveals several areas where decisions were made on limited or imprecise information. Specifically, two problems become apparent: (1) The current preclinical screening methods to identify promising therapies that should move forward into clinical testing are inadequate; (2) Tools to assist in the evaluation of early Phase I/II clinical results, identifying those therapies most likely to be successful, are lacking. Underlying these limitations has been the traditional reductionist approach to these problems, with a focus on a specific intervention leading in a linear fashion to a final outcome, independent of the intervening pathways. Systems biology, with an emphasis on understanding the intervening components and providing predictive models to anticipate these results, offers an alternate approach. In addition to the fundamental knowledge that can be created, this approach offers a vital translational research tool for the evaluation of therapies thought to have efficacy in improving outcomes following vascular intervention. Through development of a robust predictive model of vascular adaptation, linking hemodynamics with gene regulatory networks and cell dynamics, the following applications may be achieved:

- Streamlined screening of biologic compounds likely to yield a positive phenotype following vascular intervention. Coupled with high throughput genomics, this model will facilitate an evaluation of the effect of a compound on the network system, providing the ability to select promising agents that should move forward into clinical investigation.
- *In silico* experimentation using the network to identify previously unidentified targets for treatment.
- Monitor Phase I/II clinical trials to evaluate the impact of the treatment on the network "state", recognizing those compounds most appropriate for transition into large human testing.
- Clinical analysis tool to assist surgeons in patient care. Coupled with high resolution vascular imaging, the proposed model will assist physicians in the identification of those patients and regions most at risk for accelerated failure.

References

1. Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, Haase N, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell CJ, Roger V, Rumsfeld J, Sorlie P, Steinberger J, Thom T, Wasserthiel-Smoller S, Hong Y (2007) Heart disease and stroke statistics–2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Circulation 115:e69–171
2. Bhatt DL, Steg PG, Ohman EM, Hirsch AT, Ikeda Y, Mas JL, Goto S, Liau CS, Richard AJ, Rother J, Wilson PW (2006) International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. JAMA 295:180–189
3. Conte MS, Bandyk DF, Clowes AW, Moneta GL, Seely L, Lorenz TJ, Namini H, Hamdan AD, Roddy SP, Belkin M, Berceli SA, DeMasi RJ, Samson RH, Berman SS (2006) Results of PREVENT III: a multicenter, randomized trial of edifoligide for the prevention of vein graft failure in lower extremity bypass surgery. J Vasc Surg 43:742–751

4. Conte MS, Mann MJ, Simosa HF, Rhynhart KK, Mulligan RC (2002) Genetic interventions for vein bypass graft disease: a review. J Vasc Surg 36:1040–1052
5. Glagov S (1994) Intimal hyperplasia, vascular modeling, and the restenosis problem. Circulation 89:2888–2891
6. Glagov S, Bassiouny HS, Giddens DP, Zarins CK (1995) Pathobiology of plaque modeling and complication. Surg Clin North Am 75:545–556
7. Wentzel JJ, Krams R, Schuurbiers JC, Oomen JA, Kloet J, Der Giessen WJ, Serruys PW, Slager CJ (2001) Relationship between neointimal thickness and shear stress after Wallstent implantation in human coronary arteries. Circulation 103:1740–1745
8. Newby AC, Baker AH (1999) Targets for gene therapy of vein grafts. Curr Opin Cardiol 14:489–494
9. Bassiouny HS, White S, Glagov S, Choi E, Giddens DP, Zarins CK (1992) Anastomotic intimal hyperplasia: mechanical injury or flow induced. J Vasc Surg 15:708–716
10. Cambria RP, Megerman J, Brewster DC, Warnock DF, Hasson J, Abbott WM (1987) The evolution of morphologic and biomechanical changes in reversed and in-situ vein grafts. Ann Surg 205:167–174
11. Zhang WD, Bai HZ, Sawa Y, Yamakawa T, Kadoba K, Taniguchi K, Masuda J, Ogata J, Shirakura R, Matsuda H (1999) Association of smooth muscle cell phenotypic modulation with extracellular matrix alterations during neointima formation in rabbit vein grafts. J Vasc Surg 30:169–183
12. Zwolak RM, Adams MC, Clowes AW (1987) Kinetics of vein graft hyperplasia: association with tangential stress. J Vasc Surg 5:126–136
13. Mills JL, Sr., Wixon CL, James DC, Devine J, Westerband A, Hughes JD (2001) The natural history of intermediate and critical vein graft stenosis: recommendations for continued surveillance or repair. J Vasc Surg 33:273–278
14. Vesti BR, Primozich J, Bergelin RO, Strandness E Jr (2001) Follow-up of valves in saphenous vein bypass grafts with duplex ultrasonography. J Vasc Surg 33:369–374
15. DeBakey ME (1979) Research related to surgical treatment of aortic and peripheral vascular disease. Circulation 60:1619–1635
16. Fernandez CM, Goldman DR, Jiang Z, Ozaki CK, Tran-Son-Tay R, Berceli SA (2004) Impact of shear stress on early vein graft remodeling: a biomechanical analysis. Ann Biomed Eng 32:1484–1493
17. Jiang Z, Wu L, Miller BL, Goldman DR, Fernandez CM, Abouhamze ZS, Ozaki CK, Berceli SA (2004c) A novel vein graft model: adaptation to differential flow environments. Am J Physiol Heart Circ Physiol 286:H240-H245
18. Berceli SA, Jiang Z, Klingman NV, Schultz GS, Ozaki CK (2006) Early differential MMP-2 and -9 dynamics during flow-induced arterial and vein graft adaptations. J Surg Res 134: 327–334
19. Jiang Z, Berceli SA, Pfahnl CL, Wu L, Killingsworth CD, Vieira FG, Ozaki CK (2004b) Impact of IL-1beta on flow-induced outward arterial remodeling. Surgery 136:478–482
20. Jiang Z, Berceli SA, Pfahnl CL, Wu L, Goldman D, Tao M, Kagayama M, Matsukawa A, Ozaki CK (2004a) Wall shear modulation of cytokines in early vein grafts. J Vasc Surg 40:345–350
21. Jiang Z, Yu P, Tao M, Fernandez C, Ifantides C, Moloye O, Schultz GS, Ozaki CK, Berceli SA (2007b) TGF-beta- and CTGF-mediated fibroblast recruitment influences early outward vein graft remodeling. Am J Physiol Heart Circ Physiol 293:H482-H488
22. Jiang Z, Shukla A, Miller BL, Espino DR, Tao M, Berceli SA, Ozaki CK (2007a) Tumor necrosis factor-alpha and the early vein graft. J Vasc Surg 45:169–176
23. Meyerson SL, Skelly CL, Curi MA, Shakur UM, Vosicky JE, Glagov S, Schwartz LB, Christen T, Gabbiani G (2001) The effects of extremely low shear stress on cellular proliferation and neointimal thickening in the failing bypass graft. J Vasc Surg 34:90–97
24. Schwartz LB, O'Donohoe MK, Purut CM, Mikat EM, Hagen PO, McCann RL (1992) Myointimal thickening in experimental vein grafts is dependent on wall tension. J Vasc Surg 15:176–186

25. Ribba B, Marron K, Agur Z, Alarcon T, Maini PK (2005) A mathematical model of Doxorubicin treatment efficacy for non-Hodgkin's lymphoma: investigation of the current protocol through theoretical modelling results. Bull Math Biol 67:79–99
26. Ribba B, Colin T, Schnell S (2006) A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7
27. Li NY, Verdolini K, Clermont G, Mi Q, Rubinstein EN, Hebda PA, Vodovotz Y (2008) A patient-specific in silico model of inflammation and healing tested in acute vocal fold injury. PLoS ONE 3:e2789
28. Mi Q, Riviere B, Clermont G, Steed DL, Vodovotz Y (2007) Agent-based model of inflammation and wound healing: insights into diabetic foot ulcer pathology and the role of transforming growth factor-beta1. Wound Repair Regen 15:671–682
29. Loth F, Jones SA, Zarins CK, Giddens DP, Nassar RF, Glagov S, Bassiouny HS (2002) Relative contribution of wall shear stress and injury in experimental intimal thickening at PTFE end-to-side arterial anastomoses. J Biomech Eng 124:44–51
30. Yang C, Tang D, Liu SQ (2003) A multi-physics growth model with fluid-structure interactions for blood flow and re-stenosis in rat vein grafts: a growth model for blood flow and re-stenosis in grafts. Comput Struct 81:1041–1058
31. Serini G, Ambrosi D, Giraudo E, Gamba A, Preziosi L, Bussolino F (2003) Modeling the early stages of vascular network assembly. EMBO J 22:1771–1779
32. Bailey AM, Thorne BC, Peirce SM (2007) Multi-cell agent-based simulation of the microvasculature to study the dynamics of circulating inflammatory cell trafficking. Ann Biomed Eng 35:916–936
33. Thorne BC, Bailey AM, Peirce SM (2007) Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. Brief Bioinform 8:245–257
34. Facciotti MT, Bonneau R, Hood L, Baliga NS (2004) Systems biology experimental design – Considerations for building predictive gene regulatory network models for prokaryotic systems. Curr Genomics 5:527–544
35. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
36. Ehsan A, Mann MJ, Dell'Acqua G, Dzau VJ (2001) Long-term stabilization of vein graft wall architecture and prolonged resistance to experimental atherosclerosis after E2F decoy oligonucleotide gene therapy. J Thorac Cardiovasc Surg 121:714–722
37. Ehsan A, Mann MJ, Dell'Acqua G, Tamura K, Braun-Dullaeus R, Dzau VJ (2002) Endothelial healing in vein grafts: proliferative burst unimpaired by genetic therapy of neointimal disease. Circulation 105:1686–1692
38. Mann MJ, Dzau VJ (2000) Therapeutic applications of transcription factor decoy oligonucleotides. J Clin Invest 106:1071–1075
39. Mann MJ, Whittemore AD, Donaldson MC, Belkin M, Conte MS, Polak JF, Orav EJ, Ehsan A, Dell'Acqua G, Dzau VJ (1999) Ex-vivo gene therapy of human vascular bypass grafts with E2F decoy: the PREVENT single-centre, randomised, controlled trial. Lancet 354:1493–1498
40. Mann MJ, Conte MS (2003) Transcription factor decoys for the prevention of vein bypass graft failure. Am J Cardiovasc Drugs 3:79–85

Modeling and Role of Leukocytes in Inflammation

Minki Hwang, Scott A. Berceli, and Roger Tran-Son-Tay

Introduction

Inflammation is a protective response of the body to infection or injury. When the body tissue is damaged by infection or injury, inflammatory response is triggered to remove the foreign invaders or start the healing process. In some diseases, such as atherosclerosis and rheumatoid arthritis, the inflammatory response is triggered inappropriately and the inflammatory cells damage the normal tissues. This suggests that those diseases can be treated by interfering with the inflammatory process.

Inflammation is manifested by pain, elevated temperature, redness, and swelling. Early events of the inflammation, which are usually independent of whether they are triggered by infection or injury, are changes of the volume (and blood flow rate as a result) and permeability of the blood vessel in the region of inflammation. Leukocytes are recruited from the blood stream to the site of inflammation, which are facilitated by the changed permeability of the vessel wall. Recruited leukocytes kill pathogens, and remove them by phagocytosis.

When the leukocytes migrate through the vessel wall, they change their shape so that they can pass through the narrow endothelial junctions. This means that the deformability of leukocytes is an important factor in the inflammatory process. Over the past several decades, our ability to characterize and simulate leukocyte deformability and behavior has tremendously improved. Some of these leukocyte models and simulation techniques are reviewed in this chapter.

M. Hwang
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA

S.A. Berceli
Department of Surgery, University of Florida College of Medicine, Gainesville, FL, USA
and
Malcom Randall Veterans Affairs Medical Center, Gainesville, FL, USA

R. Tran-Son-Tay (✉)
Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, USA
and
Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
e-mail: rtst@ufl.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_13,

Leukocytes and Inflammation

Leukocytes and other blood cells are differentiated from the pluripotent hematopoietic stem cell in the bone marrow. Several hundred billion leukocytes are produced daily due to the short half-life which is typically hours to days [1]. There are five types of leukocytes: lymphocytes, monocytes, neutrophils, eosinophils, and basophils. Monocytes develop into macrophages when they enter the tissue from the blood vessel. Neutrophils, monocytes, macrophages, and eosinophils are phagocytes because they engulf foreign invaders such as bacteria. Basophils, eosinophils, and neutrophils are called granulocytes because granules are observed in their cytoplasm. Neutrophils are also called polymorphonuclear leukocytes because their nucleus consists of several nuclear lobes. Neutrophils, the most abundant type of leukocytes, typically kill five to twenty bacteria during their life span of several hours to few days. On the other hand, monocytes constitute only 1–6% of all leukocytes in the blood, but kill about 100 bacteria during their life span of several months [1]. Lymphocytes play an important role in the acquired immune response of the body.

On the surface of the leukocytes, there are projections of plasma membrane, called microvilli. There are 500–10,000 microvilli on a leukocyte surface [2]. The length of the microvilli ranges from 50 to 3,500 nm [3], and the surface area of the tip of a microvillus is approximately 0.01 μm^2 [4]. At the tip of the microvilli, there are various adhesion molecules which mediate cell-to-cell interactions. The microvillus elongates when force is exerted. At small force, the microvillus stretches like an elastic spring, and the spring constant has been measured to be 43 $pN/\mu m$ in case of neutrophil microvilli [5]. The leukocyte plasma membrane has numerous wrinkles and folds which accommodate surface area increase when the leukocyte deforms from its initial spherical shape. The surface area increases more than 100% when the leukocyte spreads on the endothelium before extravasation [6].

Among the adhesion molecules on the surface of the leukocytes and endothelial cells, selectin and integrin play critical roles in the leukocyte recruitment to inflammation sites. Selectins are transmembrane molecules composed of N-terminal extracellular domain, epidermal growth factor, consensus repeats, transmembrane domain, and cytoplasmic tail. L-selectin is expressed on leukocytes and binds to the ligands on endothelial cells. P-selectin is expressed on endothelium and platelets, and binds to PSGL-1 on leukocyte. E-selectin is expressed on endothelial cells and binds to ligands on leukocytes. Integrins are heterodimeric glycoproteins which are subdivided according to their subunits. Integrins are also transmembrane molecules composed of extracellular domain and cytoplasmic tail which is linked to the cytoskeleton of the cell. VLA-4 integrin on leukocyte binds to VCAM-1 ligand on endothelium upon activated by chemotactic agents.

Leukocytes migrate to inflammation sites by sensing the inflammatory signals originated at the sites in the tissue. When the circulating leukocytes migrate to those sites, they go through several steps: initial capture, rolling, firm adhesion, and transendothelial migration. Leukocytes are captured from the blood stream through the binding between the selectins expressed on the leukocytes and the

ligands on the surface of the endothelial cells. The captured leukocytes roll on the surface of the endothelium under the combined effect of shear stress imposed by the blood flow and the association/dissociation between the endothelial selectins and the leukocyte ligands. When the chemokine receptors on the rolling leukocyte detect chemokines on the endothelial cells, the leukocyte integrins which were previously low affinity state undergo conformational change to become high affinity state. Through the binding between these high affinity integrins on the leukocyte and the ligands on the endothelium, the rolling leukocytes firmly adhere to the endothelium. The firm adhesion triggers the cytoskeleton diapedesis machinery leading to transendothelial migration [7].

Leukocyte Rolling Mechanism

Leukocyte rolling before transendothelial migration in the inflammatory response is a movement coordinated by fluid shear and binding between adhesion molecules on the leukocyte and endothelium. When the bonds at the trailing edge of the contact area dissociate, the leukocyte rotates forward due to the fluid force, and the probability of bond association at the leading edge increases. Leukocyte rolling is mediated mostly by selectins, and different rolling velocities have been observed on different types of selectins [8].

It has been observed that the rolling velocity increases as fluid shear stress increases. Lawrence and Springer [9] reproduced rolling of neutrophils on artificial lipid bilayers containing purified CD62 in vitro, and observed that the rolling velocity increased with increasing shear stress. They also observed that the rolling velocity increased with decreasing selectin density. Alon et al. [10] reported that the rolling velocity of blood T lymphocytes increased when shear stress was increased in a parallel flow chamber assay. Dong and Lei [11] observed that rolling velocities of HL-60 cells, which express comparable levels of PSGL-1 (P-selectin ligand) as most human leukocytes, increased on a surface with P-selectin adsorbed as wall shear stress increased in their in vitro side-view flow assay.

An interesting phenomenon in the leukocyte rolling is the stop-and-go motion. The leukocytes frequently stop for some time as they roll on the ligands, and it presumably results from the fact that the binding between the adhesion molecules is stochastic in nature. Goetz et al. [12] measured the instantaneous velocity of bovine neutrophils interacting with lipopolysaccharide-stimulated bovine aortic endothelium at various shear stresses, and confirmed that the neutrophils translated with a nonconstant velocity. The population average variance in the instantaneous velocity was at least 2 orders of magnitude higher than the theoretical variance generated from experimental error. Smith et al. [13] obtained the dissociation constants (k_{off}) for P-, E-, L-selectin from the pause times in the neutrophil rolling in their parallel plate flow chamber. The elasticity of the microvilli on the surface of the leukocyte also has a role in the leukocyte rolling. Park et al. [14] compared the rolling characteristics of PSGL-1 coated microbead and neutrophil, and reported that the

dissociation rates for the PSGL-1 microbeads were briefer than those of neutrophils. They also observed the elongated microvilli at the rear of the rolling neutrophils using scanning electron microscope.

Fluid shear affects the number of tethered and rolling leukocytes. Finger et al. [15] reported that the number of tethered and rolling neutrophils on E-selectin and P-selectin decreased as the shear stress increased. However, they observed that, in case of L-selectin, the number of tethered and rolling neutrophils increased as the shear stress increased up to around 1 dynes/cm^2, and then, the number decreased with increasing shear stress over about 1 dynes/cm^2. Later, Lawrence et al. [16] reported that the threshold shear exists for E-selectin and P-selectin as well. Dwir et al. [17] identified short-duration (4–20 ms) tethers of L-selectin below 0.3 dynes/cm^2 of shear stress, and postulated that shear-mediated multivalent tether is required for the L-selectin tethers to become stabilized. An interesting observation was made by Chen and Springer [18] that the tether formation is related to shear rate rather than to shear stress. They showed that the tether frequency of neutrophils to P-selectin reached the maximum value at the shear rate of $\sim 100\,s^{-1}$ regardless of viscosity.

Leukocyte Rheological Models

Leukocyte deformability plays an important role in the extravasation of the leukocyte in the inflammatory response. For the investigation of the leukocyte deformability, micropipette aspiration experiments have been extensively used and reviewed [19, 20]. In those experiments, single cell is aspirated into a micropipette of which the inner diameter is less than the cell diameter. The position of the leading edge of the cell is tracked as a function of time for various suction pressures. Comparison of the experimental data with appropriate models provides the validity of the model, and thus the mechanical properties of the cells. Recovery experiments have also been used to gain insights into the cell rheological behavior. In those experiments, a single cell after being aspirated inside a micropipette is expelled from it, and the dimensions of the cell as it recovers its initial, resting, shape are tracked as a function of time [21].

Viscoelastic Solid Model

One of the earliest models describing the rheological properties of leukocytes is the viscoelastic solid model proposed by Schmid-Schönbein et al. [22]. They performed micropipette aspiration experiment on human leukocytes for small deformation (The displacement of the leukocyte tongue inside the pipette was up to $\sim 1\,\mu m$), and compared the experimental data with an analytical solution from the viscoelastic solid model. Though the model solution and the experimental data agreed well for small deformation, it is found that leukocytes behave as liquids and not as solids [23].

Cortical Shell-Liquid Core Model

When the leukocyte is aspirated into the micropipette, the leukocyte flows continuously into the pipette once the pressure difference is above a threshold value. To model the liquid-like behavior of the leukocyte, Yeung and Evans [24] developed a cortical shell-liquid core model. In their model, the viscous-liquid core is surrounded by an anisotropic viscous cortical shell with a lateral tension. Their inclusion of the cortical tension is based on the observation that the cell recovers to nearly perfect spheres after large extensions into micropipettes, and that the threshold pressure scales inversely with pipette diameter. They used a variational approach to obtain a mathematical expression of the rate of leukocyte entry into the pipette as a function of the ratio of pipette radius to instantaneous segment radius, viscosity of the cortical shell, and the projection length.

On the basis of the theoretical results of Yeung and Evans [24], Needham and Hochmuth [25] developed a linearized version of the leukocyte entry rate, and obtained the following expression for the cytoplasmic viscosity.

$$\mu = \frac{(\Delta P)\, R_{\mathrm{P}}}{(\mathrm{d}L_{\mathrm{P}}/\mathrm{d}t)\, m \left(1 - \frac{1}{\bar{R}}\right)} \tag{1}$$

where μ is cytoplasmic viscosity, ΔP is the excess aspiration pressure above the yield threshold for initial entry, R_{p} is the pipette radius, L_{p} is the length of the cell in the pipette, t is time, $\bar{R} = R/R_{\mathrm{P}}$, R is the radius of the cell outside the pipette, m is a constant proportional to the ratio of the surface dissipation in the cortex to the dissipation in the cytoplasm. They found that the complex cytoplasm inside unactivated neutrophils behaves as a nearly Newtonian fluid with a viscosity on the order of 10^2 Pa s over almost a two order of magnitude range in aspiration pressure.

The cortical shell – liquid (Newtonian) core model has also been used to describe the recovery process of leukocyte by Tran-Son-Tay et al. [21, 26]. They performed experiments in which a passive human neutrophil is deformed into an elongated shape as it is aspirated into a small glass pipette, and, after release, the neutrophil was monitored as it recovered its natural spherical shape. They obtained the parametric solutions of the recovery problem using the cortical shell – liquid drop model, and the model agreed well with the experimental data (Fig. 1). Interestingly, when the neutrophil was held in the pipette for less than 5 s, the cell recovered its initial shape rapidly when expelled, suggesting that the cell behaves as a viscoelastic liquid.

Non-Newtonian Model

Tsai et al. [27] observed that the cytoplasmic viscosity of the neutrophils decreased with increasing shear rate in their micropipette aspiration experiment, and

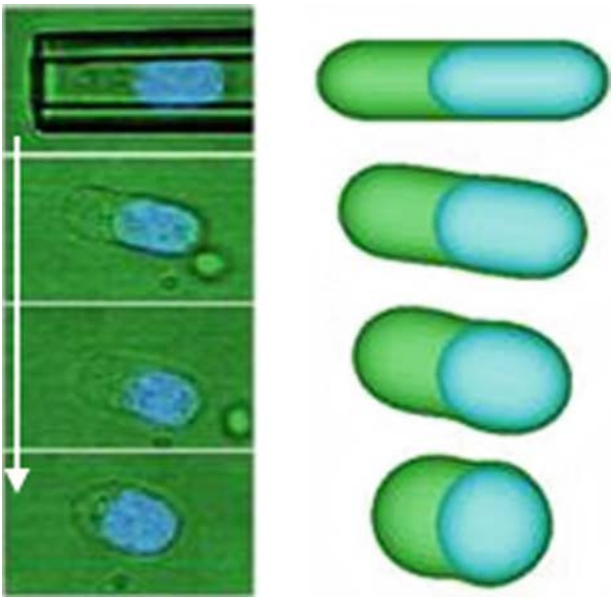


Fig. 1 Leukocyte recovery process. A neutrophil elongated in the micropipette recovers its initial spherical shape after release

they found that the following power-law relation approximates the non-Newtonian behavior of the neutrophils.

$$\mu = \mu_c \left(\frac{\dot{\gamma}_m}{\dot{\gamma}_c} \right)^{-b} \tag{2}$$

where μ is the apparent cytoplasmic viscosity, $\dot{\gamma}_m$ is the mean shear rate, μ_c is the characteristic viscosity at characteristic shear rate $\dot{\gamma}_c$, and b is a material coefficient. Their simulation using this power-law relationship predicted the rapid entry at the final phase of the aspiration, which is possibly due to shear thinning.

Compound Drop Model

Although the power-law model can describe the non-Newtonian behavior of leukocyte, it can not predict the rapid recovery after the leukocyte is held in the micropipette for a very short time. Also, the model does not take into account the complex internal structure of the cell. Kan et al. [28] proposed the compound drop model for more realistic description of the leukocyte behavior. The compound drop model consists of three layers: a thin cortical shell with a persistent isotropic surface tension, a thick layer of cytoplasm, and a nucleus. They simulated the deformation of the leukocyte by extensional flow and recovery of the same leukocyte (Fig. 2). They found that the compound drop model explains many experimental results which previous models could not explain, such as the rapid recoil of the leukocyte,

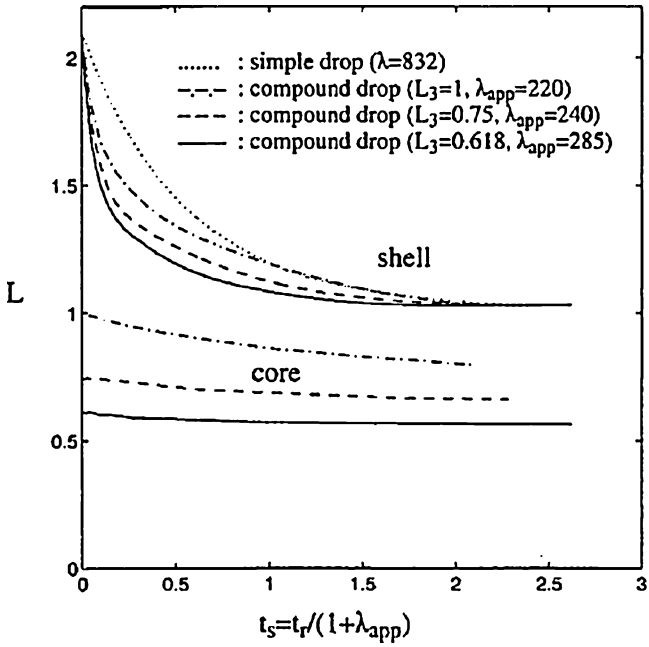


Fig. 2 Recovery simulation of compound drops with different initial core deformations. The lengths of different layers in the compound drops decrease as the drops recover their original shapes. (Taken from [28])

mostly due to the fact that the compound drop model takes into account different viscosities of the nucleus and cytoplasm-cortex complex [29]. The compound drop model also shows the behavior of Newtonian liquid drop when the time scales of the nucleus and cytoplasm-cortex complex are comparable. They also found that the flow dynamics inside the three layers affect the cell apparent viscosity.

Kan et al. [30] simulated the deformation of the compound drop model adhered on a surface in a shear flow. They solved the governing equations for Stokes flow, and used the ELAFINT (Eulerian-Lagrangian Algorithm For INterface Tracking) technique [31] to track the interfaces between the cell and the outer flow. They found that the adhered leukocyte deforms more when the initial contact angle is higher and/or the capillary number is higher. They also found that the presence of a highly viscous nucleus in the cell changes deformation time compared to the simple liquid drop without the nucleus, although the steady state shapes remain similar.

Leukocyte Rolling Models

As described in the previous sections, the leukocyte rolling is a complex process which involves many biological and mechanical factors. This section outlines the leukocyte rolling models proposed so far.

Bond Model

The performance of the rolling model depends on the biological and mechanical component models, and the model of bonds between adhesion molecules on the surface of the leukocyte and endothelium is an important component model. The most commonly used bond model is spring model [32] in which the force in the bond is proportional to the elongation of the bond. The forward and reverse reaction rates between receptor and ligand are given by:

$$k_{\mathrm{f}} = k_{\mathrm{f0}} \exp\left(-\frac{\sigma_{\mathrm{ts}} \left(x_{\mathrm{m}} - \lambda\right)^2}{2k_{\mathrm{b}}T}\right) \tag{3}$$

$$k_{\mathrm{r}} = k_{\mathrm{r0}} \exp\left(\frac{\left(\sigma - \sigma_{\mathrm{ts}}\right) \left(x_{\mathrm{m}} - \lambda\right)^2}{2k_{\mathrm{b}}T}\right) \tag{4}$$

where k_{f} is forward reaction rate, k_{r} is reverse reaction rate, k_{f0} is initial forward reaction rate, k_{r0} is initial reverse reaction rate, σ is spring constant, σ_{ts} is transition spring constant, x_{m} is the length of the bond, λ is equilibrium bond length, k_{b} is Boltzmann constant, and T is the temperature.

Equation (3) states that the rate of bond association decreases as the distance between receptor and ligand increases. In (4), the bond dissociation rate increases with increasing distance between receptor and ligand when σ–σ_{ts} is positive (slip bond), and the bond dissociation rate decreases with increasing distance between receptor and ligand when σ–σ_{ts} is negative (catch bond).

Rigid Sphere Model

Hammer and Apte [33] modeled the leukocyte as a rigid sphere with microvilli on the leukocyte surface. They used the bond model described in the previous section for the reaction between the adhesion molecules on the tip of the microvilli and on the surface on which the leukocyte is rolling. The hydrodynamic force was calculated based on Stokes' solution [34, 42]. Their model enabled them to examine the effects of microvilli density on cell surface, receptor density on microvillus tip, the stiffness of the receptor-ligand bond, and the magnitude of the hydrodynamic force. The simulation results showed the velocity fluctuation which is one of the characteristics of leukocyte rolling.

Compound Drop Model

N'Dri et al. [35] simulated the rolling of deformable compound drop model which has nucleus inside the cell (Fig. 3). They solved the Navier-Stokes equations to

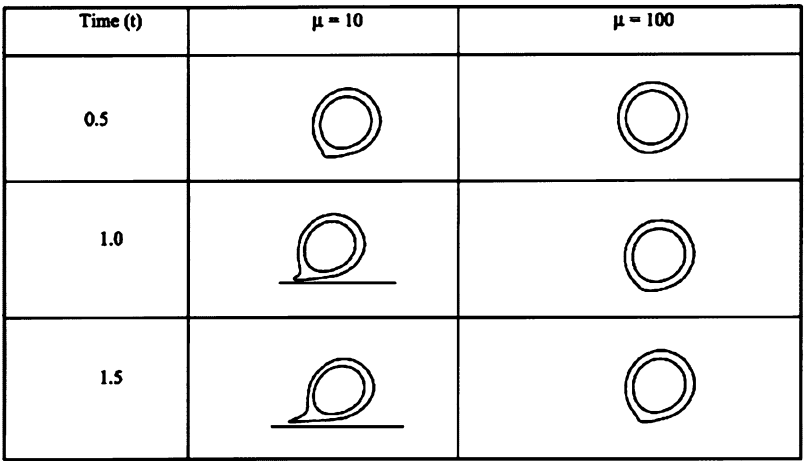


Fig. 3 Simulation of compound drop rolling. The cell with lower cytoplasmic viscosity deforms more. (Taken from [35])

simulate the flow field, and used immersed boundary technique [36] to incorporate the effect of the cell in the simulation. The Eulerian-Lagrangian technique [37, 38] was used to track the moving interfaces between the outer flow and the cell, and between the cytoplasm and nucleus. They simulated the association and dissociation of the adhesion molecules on the leukocyte and the surface using the bond model described in the previous section. They found that the presence of nucleus inside the cell increases the bond life time, and decreases the cell rolling velocity. They also found that a bigger cell rolls faster, and causes bond lifetime decrease, possibly due to the fact that a larger cell experiences greater hydrodynamic force from the blood stream.

Deformable 3-Dimensional Model

Jadhav et al. [39] extended the model of [35] to 3-dimensional model. They modeled the cell as an elastic capsule, and used the immersed boundary method to simulate the motion of the capsule in a shear flow. The nucleus was not included in the cell model. They simulated the forward and reverse reactions between the adhesion molecules on the leukocyte and the surface based on the probabilities of each reaction [33]. They showed the rolling velocity fluctuation for 0.7 s at 200 s^{-1} of shear rate. They observed that higher membrane stiffness made the cell roll faster, and the number of bonds decreased with increasing membrane stiffness. Bond lifetime also decreased with increasing membrane stiffness.

Pappu and Bagchi [40] examined a similar model as that of Jadhav et al. [39]. They showed the forward movement of the contact area between the cell and the

surface, and the displacement of the cell as a function of time for up to 3 s, which showed the stop-and-go motion. They observed a sideway motion during the rolling due to discrete nature of microvilli distribution. They also found that most of the adhesion force is sustained by only 1–3 tethered microvilli in the rear-most part of a cell.

Systems Biology Approach

Tang et al. [41] developed an in silico model of leukocyte rolling, activation, and adhesion. They modeled the leukocyte as having rectangular contact area which is discretized such that each element has adhesion molecules. The surface on which the leukocyte rolls is also discretized and adhesion molecules are contained in the elements. The bindings between the adhesion molecules were simulated using the probabilities of bond formation/dissociation. They applied a fluid force which affects the dissociation of the bonds, and when all the bonds have been dissociated at the rear row of the contact area, the leukocyte rolls forward. Their simulation results of rolling behavior agreed well with in vitro experimental data, and the transition from rolling to adhesion was also simulated.

Summary and Conclusion

Over the past several decades, a large number of biologically and rheologically interesting phenomena of leukocyte behaviors have been discovered. Many rheological models of leukocyte from simple solid model to compound drop model have been developed to account for the experimentally observed phenomena. However, these models apply to a limited range of experimental conditions or explain only a part of the experimentally observed rheological phenomena. A more accurate rheological model is needed to account for all the phenomena from small to large deformation and flow characteristics. This is also true for leukocyte rolling models. Although numerous leukocyte rolling models have been developed to elucidate the rolling mechanism of leukocyte, a complete model of leukocyte dynamics is still needed to describe the whole process from capture by endothelium to extravasation. This complete model will require the usage of a large amount of computational resources due to the fact that the whole process of leukocyte trafficking is a complex process. The advancement of computing technology is needed in this regard. Recently, systems biology approach is becoming popular in almost all areas of biology and medicine. Although the application of the systems biology approach in the study of leukocyte dynamics is in its beginning stage, systems biology approach coupled with mathematical/mechanical models is a promising methodology in that it can incorporate all the experimentally observed phenomena while benefiting from the theory of mechanics.

References

1. Silverthorn DU, Ober WC, Garrison CW et al (2001) Human physiology an integrated approach. Prentice Hall, Upper Saddle River
2. Knutton S, Sumner MCB, Pasternak CA (1975) Role of microvilli in surface changes of synchronized P815Y mastocytoma cells. J Cell Biol 66:568–576
3. Bruehl RE, Springer TA, Bainton DF (1996) Quantitation of L-selectin distribution on human leukocyte microvilli by immunogold labeling and electron microscopy. J Histochem Cytochem 44:835–844
4. Bongrand P, Bell G I (1984) Cell-cell adhesion: parameters and possible mechanisms. In: Perelson AS, DeLisi C, Wiegel FW (eds) Cell surface dynamics, Concepts and models. Marcel Dekker, New York
5. Shao J-Y, Ting-Beall HP, Hochmuth RM (1998) Static and dynamic lengths of neutrophil microvilli. Proc Natl Acad Sci U S A 95:6797–6802
6. Dewitt S, Hallett M (2007) Leukocyte membrane "expansion": a central mechanism for leukocyte extravasation. J Leukocyte Biol 81:1160–1164
7. Shuhaiber JH, Evans AN, Massad MG et al (2002) Mechanisms and future directions for prevention of vein graft failure in coronary bypass surgery. Eur J Cardiothorac Surg 22:387–396
8. Jung U, Bullard DC, Tedder TF et al (1996) Velocity differences between L- and P-selectin-dependent neutrophil rolling in venules of mouse cremaster muscle in vivo. Am J Physiol Heart Circ Physiol 271:H2740–H2747
9. Lawrence MB, Springer TA (1991) Leukocytes roll on a selectin at physiologic flow rates: distinction from and prerequisite for adhesion through integrins. Cell 65:859–873
10. Alon R, Kassner PD, Carr MW et al (1995) The integrin VLA-4 supports tethering and rolling in flow on VCAM-1. J Cell Biol 128:1243–1253
11. Dong C, Lei XX (2000) Biomechanics of cell rolling: shear flow, cell-surface adhesion, and cell deformability. J Biomech 33:35–43
12. Goetz DJ, El-Sabban ME, Pauli BU et al (1994) Dynamics of neutrophil rolling over stimulated endothelium in vitro. Biophys J 66:2202–2209
13. Smith MJ, Berg EL, Lawrence MB (1999) A direct comparison of selectin-mediated transient, adhesive events using high temporal resolution. Biophys J 77:3371–3383
14. Park EYH, Smith MJ, Stropp ES et al (2002) Comparison of PSGL-1 microbead and neutrophil rolling: microvillus elongation stabilizes P-selectin bond clusters. Biophys J 82:1835–1847
15. Finger EB, Puri KD, Alon R et al (1996) Adhesion through L-selectin requires a threshold hydrodynamic shear. Nature 379:266–269
16. Lawrence MB, Kansas GS, Kunkel EJ et al (1997) Threshold levels of fluid shear promote leukocyte adhesion through selectin (CD62L, P, E). J Cell Biol 136:717–727
17. Dwir O, Solomon A, Mangan S et al (2003) Avidity enhancement of L-selectin bonds by flow: shear-promoted rotation of leukocytes turn labile bonds into functional tethers. J Cell Biol 163:649–659
18. Chen S, Springer TA (2001) Selectin receptor-ligand bonds: formation limited by shear rate and dissociation governed by the Bell model. Proc Natl Acad Sci U S A 98:950–955
19. Hochmuth RM (2000) Micropipette aspiration of living cells. J Biomech 33:15–22
20. Tran-Son-Tay R, Nash GB (2007) Mechanical properties of leukocytes and their effects on the circulation. In: Baskurt OK, Hardeman MR, Rampling MW et al (eds) Handbook of hemorheology and hemodynamics. IOS Press, Amsterdam, Netherlands
21. Tran-Son-Tay R, Needham D, Yeung A et al (1991) Time-dependent recovery of passive neutrophils after large deformation. Biophys J 60:856–866
22. Schmid-Schönbein GW, Sung K-LP, Tözeren H et al (1981) Passive mechanical properties of human leukocytes. Biophys J 36:243–256
23. Evans E, Yeung A (1989) Apparent viscosity and cortical tension of blood granulocytes determined by micropipet aspiration. Biophys J 56:151–160
24. Yeung A, Evans E (1989) Cortical shell-liquid core model for passive flow of liquid-like spherical cells into micropipets. Biophys J 56:139–149

25. Needham D, Hochmuth RM (1990) Rapid flow of passive neutrophils into a 4 μm pipet and measurement of cytoplasmic viscosity. J Biomech Eng 112:269–276
26. Tran-Son-Tay R, Kan H-C, Udaykumar HS et al (1998) Rheological modelling of leukocytes. Med Biol Eng Comput 36:246–250
27. Tsai MA, Frank RS, Waugh RE (1993) Passive mechanical behavior of human neutrophils: power-law fluid. Biophys J 65:2078–2088
28. Kan H-C, Udaykumar HS, Shyy W et al (1998) Hydrodynamics of a compound drop with application to leukocyte modeling. Phys Fluids 10:760–774
29. Kan H-C, Shyy W, Udaykumar HS et al (1999a) Effects of nucleus on leukocyte recovery. Ann Biomed Eng 27:648–655
30. Kan H-C, Udaykumar HS, Shyy W et al (1999b) Numerical analysis of the deformation of an adherent drop under shear flow. J Biomech Eng 121:160–169
31. Shyy W, Udaykumar HS, Rao MM et al (1996) Computational fluid dynamics with moving boundaries. Taylor and Francis, Philadelphia
32. Dembo M, Torney DC, Saxman K et al (1988) The reaction-limited kinetics of membrane-to-surface adhesion and detachment. Proc R Soc Lond Ser B 234:55–83
33. Hammer DA, Apte SM (1992) Simulation of cell rolling and adhesion on surfaces in shear flow: general results and analysis of selectin-mediated neutrophil adhesion. Biophys J 63:35–57
34. Goldman AJ, Cox RG, Brenner H (1967a) Slow viscous motion of a sphere parallel to a plane wall. I. Motion through a quiescent fluid. Chem Eng Sci 22:637–652
35. N'Dri NA, Shyy W, Tran-Son-Tay R (2003) Computational modeling of cell adhesion and movement using a continuum-kinetics approach. Biophys J 85:2273–2286
36. Peskin CS (1977) Numerical analysis of blood flow in the heart. J Comput Phys 25:220–252
37. Shyy W, Kan H-C, Udaykumar HS et al (1999) Interaction between fluid flows and flexible structures. In: Shyy W, Narayanan R (eds) Fluid dynamics at interfaces. Cambridge University Press, Cambridge, UK
38. Shyy W, Francois M, Udaykumar HS et al (2001) Moving boundaries in micro-scale biofluid dynamics. Appl Mech Rev 5:405–453
39. Jadhav S, Eggleton CD, Konstantopoulos K (2005) A 3-D computational model predicts that cell deformation affects selectin-mediated leukocyte rolling. Biophys J 88:96–104
40. Pappu V, Bagchi P (2008) 3D computational modeling and simulation of leukocyte rolling adhesion and deformation. Comput Biol Med 38:738–753
41. Tang J, Ley KF, Hunt CA (2007) Dynamics of in silico leukocyte rolling, activation, and adhesion. BMC Syst Biol 1:14
42. Goldman AJ, Cox RG, Brenner H (1967b) Slow viscous motion of a sphere parallel to a plane wall. II. Couette Flow. Chem Eng Sci 22:653–659

Multi-modality Imaging for the Simulation of Cerebral Aneurysm Blood Flow Dynamics

Christof Karmonik, Yi Jonathan Zhang, and Robert G. Grossman

Cerebral Aneurysms

Cerebral aneurysms are pathological local expansions of intracranial arteries. These expansions can affect only a portion of the artery wall (sidewall aneurysms), the whole wall (fusiform aneurysm) or can occur at the bifurcation of arteries (bifurcation aneurysm, Fig. 1).

The size of these aneurysms can vary ranging from a few millimeters (small) to centimeters (large). While symptoms of cerebral aneurysms may include ischemic events, seizures, headaches or cranial nerve palsy, sometimes, aneurysms may not show any symptoms at all. The real danger of a cerebral aneurysm is rupture, an event, in which blood exits the aneurysm wall into the brain parenchyma, resulting in subarachnoid hemorrhage (SAH). SAH is the reason of about 8% of all hemorrhagic strokes (the third leading cause of death in the United States), and 4/5 of all SAH occurrences are caused by the rupture of a cerebral aneurysm [1, 2].

While exact numbers for the prevalence of cerebral aneurysms are not available, 3–6% of the adult population have been estimated to harbor such a vascular pathology [3]. This number at first appears relatively low, but aneurysm rupture is a devastating event as it can occur at any age during adulthood, most likely at or after the relatively young age of 50 affecting otherwise healthy and productive individuals. SAH, despite recent advances of medical treatment [4, 5], is a clinically catastrophic event with high morbidity (half of all people affected show significant cognitive impairment [6, 7]) and mortality, which can reach 75% [4, 8, 9].

Etiology for cerebral aneurysms is unknown; possible risk factors include:

- Smoking
- Hypertension
- Alcohol consumption
- Cocaine and amphetamine abuse
- Oral conceptive
- Plasma cholesterol concentration in the highest tertile.

C. Karmonik (✉), Y.J. Zhang, and R.G. Grossman
The Methodist Hospital Neurological Institute, Houston, TX, USA
e-mail: CKarmonik@tmhs.org

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_14,

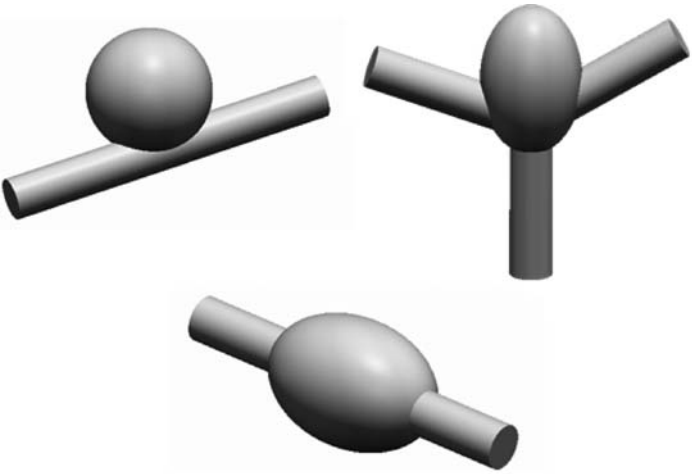


Fig. 1 Schematic illustration of different kinds of cerebral aneurysms: *Top left:* Sidewall (or saccular aneurysm), *top right:* bifurcation (or terminal aneurysm), *bottom:* fusiform aneurysm

Inherited risk factors include:

- Adult polycystic kidney disease (ADPKD)
- Type IV Ehlers–Danlos syndrome
- Hereditary hemorrhagic telangiestasia
- Neurofibromatosis Type I
- A1−antitrypsin deficiency.

Treatment of Cerebral Aneurysms

Technical advances in medical imaging (in particular 64 slice CT and 3T magnetic resonance imaging (MRI)) allow for the routine acquisition of images of the brain with an isotropic spatial resolution of less than 1 mm. In combination with intravenous contrast injections to enhance image contrast between flowing blood and stationary tissue (applied in techniques such as computed tomographic angiography (CTA), or contrast-enhances magnetic resonance imaging (ce-MRA)), unruptured cerebral aneurysms are now incidentally discovered at a higher rate than in previous years. Because of the devastating consequences of SAH, preventive intervention is often indicated compared with conservative medical management [3, 5, 10–15]. With conservative management, the morbidity is 0.1–1% per decade and mortality, 0.5–5% per decade.

Two approaches for surgical removal of a cerebral aneurysm exist: traditional open brain surgery, and more recently introduced, endovascular treatment (EVT), which is beginning to replace open surgical techniques [16–25]. In EVT, the aneurysm is occluded by metallic coils or other material (such a glue) that is placed

into the aneurysm dome directly through a catheter inserted in the femoral artery and advanced into the parent artery of the aneurysm. Mortality and morbidity rates for open surgery are 2.6–3.8% per decade and 10.9–12.1% per decade, respectively. These numbers have to be compared to 1.1% (morbidity) and 3.7–4.0 % (mortality, totally occluded) and 22–48% (mortality, partially occluded) for EVT. While it appears that EVT compares more favorably if total occlusion is achieved, large multi-center trials are still ongoing for further investigation.

To improve EVT, the development of new endovascular devices is an active area of research. Metallic coils coded with hydrogel (that expands when exposed to blood), aneurysmal neck-bridge devices, coils coded with radioactive alloys or drugs promoting fibroblast growth, or bioabsorbable polymers are only a few examples. Despite all these efforts, EVT is still not able to achieve complete occlusion in all cases, and partially occluded aneurysms are seen in many cases. In some cases, re-growth of the aneurysm occurs if only partially occluded (recurrence). One factor contributing to recurrence is thought to be hemodynamic forces. These forces are still poorly understood, and a better understanding of the aneurysmal hemodynamics is warranted to better understand the recurrence of cerebral aneurysms treated with EVT.

Aneurysm Imaging

The gold standard to confirm the existence of a cerebral aneurysm is digital subtraction angiography (DSA). With this technique, a catheter is advanced through the arterial tree to the region of interest, and an iodine contrast agent is injected into the blood stream. X-ray projection images are acquired before and after release of the contrast and subtracted. The resulting image ideally only shows the arteries wherein contrast is present (Fig. 2).

With this imaging method, a high spatial resolution in the sub-millimeter range is obtainable. The medical imager is usually called “C-arm,” because of the shape of the support structure holding the X-ray source (above the head) and the X-ray detector (below the head). This C-arm can be oriented at different angles to obtain the best projection view of the aneurysm. Many clinical systems, in particular the ones that are used during the EVT treatment of a cerebral aneurysm, are equipped with two C-arms, i.e., two source-detector pairs that can be oriented in an (almost) arbitrary angle to each other. Most often, they are oriented perpendicular, so that the interventionalist can follow the advancement of the catheter in a 3D fashion in the two projection planes. Another acquisition technique rotates one source-detector pair around the head traversing an angle range exceeding 180°. In a similar fashion to the computed tomographic (CT) image acquisitions, a 3D image can then be reconstructed from the projection images by projection reconstruction. This 3D DSA technique yield true three-dimensional spatial resolution of less than 1 mm. Recent advances in flat solid state detector technology have increased the contrast range of Hounsfield units so that soft-tissue images of the brain parenchyma can now be acquired without contrast injection. In combination with intra-arterial contrast,

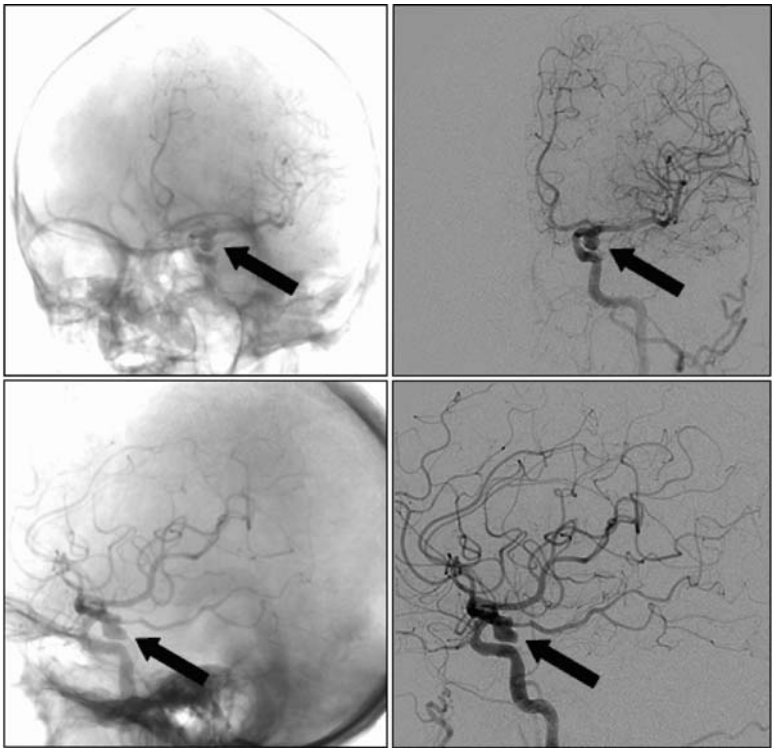


Fig. 2 DSA projection images in anterior–posterior direction (*top*) and left-right direction (*bottom*). Native images are shown on *left* and subtracted images on *right*. Aneurysm is depicted by arrow

a cerebral aneurysm can now be visualized together with its peri-aneurysmal environment. A recent report has found that geometrical details in the aneurysm shape of basilar aneurysms are better visualized by this technique than 3D DSA alone [26] (Fig. 3).

Another imaging modality that has been used to visualize cerebral aneurysms is magnetic resonance angiography (MRA). In this technique, the contrast of flowing blood compared to the stationary brain tissue is exploited to visualize cerebral arteries. In a variation, an gadolinium-based contrast-agent is injected intravenously to alter the magnetic properties of the blood, thereby enhancing blood-tissue contrast. A lower sensitivity (69–100%) and specificity (75–100%) has been found of MRA compared to DSA for the detection of at least one aneurysm in patients harboring one or more aneurysms. In addition to anatomical information, magnetic resonance imaging can also provide information about physiological flow effects. Magnetic resonance phase contrast angiography (pcMRA) is a method that creates images where the signal intensity is derived from flow effects (or moving magnetization) by applying flow-encoding gradients as part of the MRI pulse sequence. These gradients translate the velocity of the moving spins (e.g., the blood) into phase values of the magnetization which can then be visualized in the MRI phase images.

Fig. 3 3D Surface representation of the aneurysm in Fig. 2 derived from 3D DSA image data

As MRI images also contain other contributions to their phase (e.g., from inhomogeneities of the main magnetic field), two sets of complex images data are acquired: in the first set, the flow gradients are turned off, and this first set is subtracted from the second set that is sensitized to flow to produce the images. This subtraction process accentuates flow while suppressing unwanted contributions to the phase of the magnetization. Two different subtraction methods are commonly used: complex difference and phase difference. The latter is used to quantify volumetric flow rate through a vessel. To map the phase difference to a velocity, the operator has to prescribe an additional parameter, commonly called VENC (for velocity encoding). If VENC is too small, flow aliasing results may make the images unsuitable for flow analysis, and if VENC is too large, sensitivity is too poor resulting in poor contrast to noise ratio (CNR) and large errors in the flow analysis. For normal blood circulation, the maximum velocities in the major cerebral vessels are tabulated, and these values can then be used to adjust the VENC. In pathologies such as stenoses, aneurysms, and arterio-vascular malformations (AVM), however, maximum velocities can be increased several fold, which makes it more difficult to select an optimum VENC value. For flow quantification, the pcMRA slice has to be prescribed perpendicular to the vessel under investigation. The volumetric flow rate can then be calculated by the following procedure: for each voxel of the vessel cross section (determined usually by manual or computer-assisted outlining), the flow velocity for this voxel is multiplied by the voxel area followed by a summation over all voxels in this cross section. Several images are acquired at different time points during one cardiac cycle, and the average flow rate for one cardiac cycle is the average of

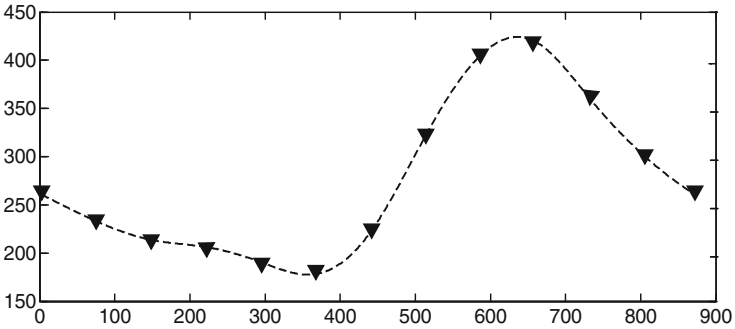


Fig. 4 Volumetric waveform measured for the aneurysm shown in Fig. 3 immediately proximal to the ostium using 2D pcMRA. Units are ml/s (y axis) and ms (x axis)

the flow rate for these different time points. Compared to the straight vessels of the peripheral vasculature, the placement of the pcMRA slice in the cerebral vessels is complicated by the tortuous nature of these vessels. The NOVA system (VasSol, Inc., www.vassolinc.com), which uses the technique of phase contrast quantitative magnetic resonance angiography (QMRA), has been developed to overcome this obstacle: based on a 3D surface reconstruction of the cerebral vasculature derived from a time-of-flight localizer, NOVA calculates the orientation of the MRI slice at a location on the vessel chosen by the operator. This orientation can then be entered on the MRI scanner console, and the pcMRA images are acquired. After acquisition, the images are transferred to the NOVA computer, and a computer-assisted analysis is performed, yielding the average flow wave form for one cardiac cycle, the volumetric flow rate time-averaged over one cardiac cycle and the blood velocities interpolated from the measurement data for typically 12–20 time points during one cardiac cycle (Fig. 4).

Computed tomography angiography (CTA) utilize the same contrast mechanism as 3D DSA, but contrast injection is performed intravenously. The spatial resolution of CTA is lower than 3D DSA (in particular the slice thickness of the reconstructed axial images is in the mm range or above), and the detection rate of cerebral aneurysms with CTA had been found to be 85–89%.

Aneurysm Rupture

At present, it is not possible to predict with certainty, if an individual asymptomatic aneurysm will rupture or remain asymptomatic. However, certain risk factors are known to be predictive of rupture. This include [27–29]:

- Size: a diameter >7 mm was reported to have an RR of 2.24 (relative risk) and a diameter >10 mm an RR of 11.6 of rupture compared with smaller aneurysms (RR 13.8 for basilar tip and RR 13.6 for vertebrobasilar locations)

- Location: An overall rupture risk of 0.25% of aneurysms located in the anterior circulation and of 1.75% for aneurysms located in the posterior circulation had been reported for aneurysms <12 mm in size in the ISUIA 2 study [18]
- Presence of a daughter aneurysm
- History of other ruptured lesions

Geometrical parameters other than size have been investigated [30], such as neck width, dome-neck ratio, height, ratio of diameter to height, ratio of neck size to aneurysm dome diameter, aspect ratio, bottle neck factor, relationship between ostium area and aneurysms volume [31–35], parent artery curvature and orientation of branching neighbored arteries [36]. All these parameters have proven unsuccessful to date to reliably predict aneurysm rupture.

Equally or even more important than geometrical parameters might be hemodynamic parameters, in particular the forces the flowing blood exerts into the aneurysm wall.

Aneurysm Hemodynamics

Hemodynamic parameters have long been thought to be of importance for aneurysm formation and rupture [37–47]. These parameters include flow velocities, dynamic pressures, static pressures, and wall shear stresses. Endothelial cells are capable of sensing biomechanical changes in their environment. They respond to shearing forces caused by time-varying flow possibly important in atherogenesis and they may also play a significant role in the development, the enlargement, and rupture of intracranial aneurysms.

Clinical imaging modalities, cannot reliably in vivo measure many of the hemodynamic parameters mentioned earlier. Insight into the hemodynamics in cerebral aneurysms has been so far mainly been limited to research studies [48–55].

Patient-Specific Computational Fluid Dynamics

To conduct patient-specific simulations of blood flow, information about vessel geometry and the flow waveform, or even better, the velocity profile should be available. Blood flow can be described in good approximation by the Navier–Stokes differential equations. As a closed analytical solution for these equations can only be derived for simple geometries, computational approximations of its solution for complex, patient-specific geometries are therefore needed. To that purpose, the 3D geometry of a cerebral aneurysm and its neighboring arteries is discretized into a 3D mesh consisting of a large number of smaller volume elements (Fig. 5).

For these small elements, it is then possible to approximate the Navier–Stokes equations and solve them at the nodes connecting the elements. Traditionally, tetrahedral elements are used, but recently, polyhedral elements have been shown to

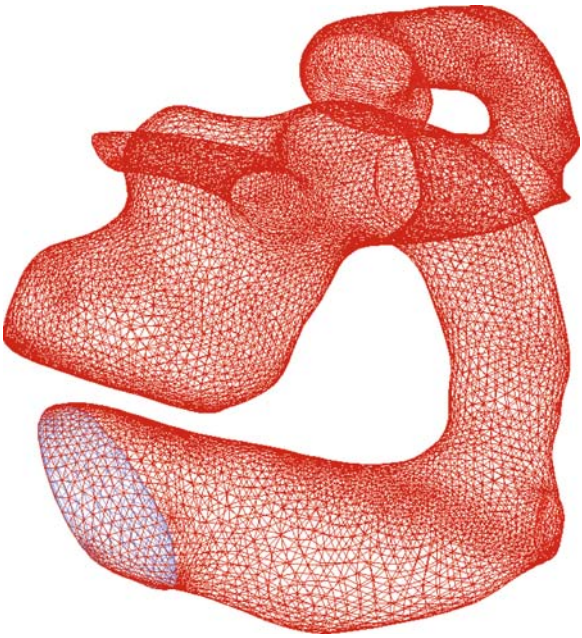


Fig. 5 Computational tetrahedral mesh derived from the 3D DSA image data shown in Fig. 3 (ANSYS GAMBIT mesh software)

possess superior properties in terms of accuracy and convergence behavior. Of importance is the elimination of the results obtained by the simulations on the characteristics of a particular mesh. To that purpose, mesh-independence calculations are usually performed in which the number of volume elements is varied (usually increased) until a convergence behavior of the results is observed. The primary hemodynamic parameter derived from these simulations is the velocity field and the pressure within the aneurysm geometry and adjacent arteries. From this field, other hemodynamic parameters, notably the wall shear stress magnitude (WSS, i.e., the force exerted by the blood tangential to the wall) and the dynamic pressure (i.e., the force perpendicular to the wall) can be extracted. While patient-specific geometries can be derived from 3D MRI or CT image datasets, of equal if not greater importance are the inflow and outflow conditions. For aneurysms in the distal cerebral circulation, the outflow boundary condition often is chosen as a zero pressure outlet, justified by the fact that there is no flow resistance in the cerebral circulation. Inflow volumetric waveforms have been either obtained from idealized waveform as averages from individual measurements of volunteers, i.e., not patient-specific, or have been measured by 2D pcMRA. Using CFD, hemodynamic parameters not accessible by current in vivo imaging methods, like the WSS or dynamic pressure, can be calculated. In addition, CFD allows to alter model parameters such as inflow rates to study the effect of derived quantities such as the WSS, a procedure that is

not feasible in a clinical setting and much more costly in experimental flow models. CFD methods have evolved to a state that they now may be considered complementary to imaging. Results derived from patient-specific flow modeling may also soon be used in therapeutic decision making.

Application of CFD for the Study of Cerebral Aneurysms

Patient-specific CFD simulations have been in applied recently in a number of studies. A recent CFD study investigated the hemodynamics in 20 cerebral aneurysms [56]. Three of these aneurysms were ruptured. Geometrical data was obtained from diagnostic three-dimensional computed tomography (CT) angiography. All aneurysms were saccular aneurysms originating at the first major bifurcation of the middle cerebral artery (MCA). The aneurysms and 20 mm of vessel data surrounding the aneurysms were used for this study. Computational meshes were generated for with about 60,000 hexagonal elements. Using finite-element techniques, blood was assumed to be an incompressible isothermal Newtonian fluid with a density of $1,000\,\mathrm{kg/m^3}$ and a viscosity of $4.0 \times 10^{-3}\,\mathrm{N/m^2\,s}$. Viscoelastic properties of the vessel wall were neglected and a rigid wall with no-slip condition was applied. Inlet flow was modeled with a pulsatile Womersley velocity profile and a traction free boundary condition was used for the outlets. Wall shear stress (WSS) distributions were calculated from the 3D velocity field data. The results of this study revealed low wall shear stresses at the aneurysm wall and alternating magnitudes of these wall shear stresses during the cardiac cycle possibly responsible for wall weakening and finally wall rupture. The average wall shear stress values at peak systole for the vessel region ($3.64 \pm 1.25\,\mathrm{N/m^2}$) were found significantly higher ($P < 0.05$) than those for the aneurysm region ($1.64 \pm 1.16\,\mathrm{N/m^2}$). The aspect ratio had a significant negative correlation ($r = -0.67$, $P < 0.05$) with the average WSS of the aneurysm region at peak systole. The spatially averaged WSS of the aneurysm region at peak systole was found to be significantly higher for the ruptured than the unruptured cases ($2.92\,\mathrm{N/m^2}$ versus $1.48\,\mathrm{N/m^2}$, $P < 0.05$). For the ruptured cases, adjacent areas of low and high WSS values were more prominent in the aneurysms area than for the unruptured cases with very low values ($\mathrm{WSS} < 0.05\,\mathrm{N/m^2}$) at the tip. Endothelial cells convert the WSS to biological signals via mechanoreceptors which leads to a modulation of gene expressions and the cellular functions of the vessel wall as response. A WSS of about $2.0\,\mathrm{N/m^2}$ is assumed to be suitable for maintaining the structure of arterial vessels and a WSS lower than $1.5\,\mathrm{N/m^2}$ will degenerate endothelial cells via the apoptotic cell cycle [57]. This study indicates that low WSS may be one of the main factors underlying the degeneration and the structural fragility of the aneurysmal wall.

Another study investigated the influence of the geometry of the proximal segment of the parent artery on the inflow into saccular cerebral aneurysm and demonstrated intensified flow impingement on the distal side of the neck with increasing arterial

curvature [36]. Blood flow inertia (centrifugal effects) rather than viscous diffusion was the predominant force driving the blood into the aneurysm sac on the curved arteries. The authors postulate that lateral saccular aneurysms located on more curved arteries are subject to higher hemodynamic stresses and recommend protection of the distal side of the aneurysm from flow impingement.

CFD methods have been applied to study flow effects in a giant vertebrobasilar aneurysm prior to treatment and have been compared to balloon test occlusion results [58]. Right vertebral artery occlusion was found preferable as it resulted in a higher contrast material washout time and a longer contrast material stasis in the aneurysm providing better conditions for aneurysm thrombosis. While the CFD calculations yielded the same results regarding the flow pattern, they also revealed a localized high pressure zone at the aneurysm after right vertebral artery conclusion adjacent to the inflow from the left vertebral artery. At 6-month follow-up, aneurysm growth was observed at this location. This case study provides an example of possible future applications of CFD in patients with cerebral aneurysm before therapeutic intervention.

A case study on a fusiform basilar aneurysm emphasizes the advantage to know the flow patterns in this kind of aneurysm prior to treatment [59]. The aneurysm model was reconstructed from image data acquired during contrast-enhanced MR angiography, and flow information was obtained using phase-contrast MR imaging methods. The study revealed a strong asymmetric in the inflow into the aneurysm with most of the flow coming from the left vertebral artery and high WSS values at the adjacent wall of the basilar artery. In a case of a giant carotid-artery aneurysm, rotational angiographic image data was used to create the computational mesh for the fluid dynamics calculations, and good agreement was found between the results of these computations and the influx dynamics of the contrast agent observed under X-ray fluoroscopy [60].

A CFD study combined with in vivo measurements of the blood inflow velocity profile using magnetic resonance phase contrast imaging [59] indicated that low wall shear stress on the aneurysm wall is correlated with aneurysm growth. Two patients with similar basilar fusiform aneurysms were followed for a 2-year period. While one aneurysm in one patient grew significantly, the similar-sized aneurysm in the other patient remained unchanged. Largest aneurysmal growth (of about 3 mm/year) was found at the location of very low wall shear stress ($<0.1\ \mathrm{N/m^2}$).

3D digital subtraction angiograms of two intracranial aneurysms were imported to a workstation for reconstruction and post-processing into 3D a format allowing delineation of (1) the extent to which the diameter of the parent artery is incorporated into the aneurysm ostium (percentage of parent artery involvement or neck angle) and (2) the aneurysm volume and surface area [31].

In a retrospective study, two subjects with AComA aneurysms were identified who received a digital subtraction angiographic (DSA) exam including 3D DSA acquisitions of separate right and left unilateral contrast injections and of one bilateral (simultaneous) injection [61]. No significant difference in aneurysm volume obtained with the bilateral injection compared to aneurysm volume of the

merged unilateral injections was noted in both cases. Pathlines computed from the CFD results were found to agree qualitatively with the results of the 3D DSA acquisitions.

Three dimensional digital subtraction angiography (3D DSA) image data of three basilar tip aneurysms were retrospectively obtained from imaging studies performed as a part of diagnostic angiograms [62]. Three wall areas of interest were defined as follows: the aneurysm area (AA) as the wall area of the basilar tip (including the aneurysm), the basilar artery area (BAA) as the vessel wall area of a healthy section of the basilar artery, and the aneurysm dome area (DA) located at the top of the aneurysm dome.

Variations in the inflow conditions result in changes of the relative average WSS magnitude for the wall areas AA and DA, with the relative average of the WSS for AA varying from 95 to 84% for case 1; from 80 to 68% for case 2 and from 86 to 76% for case 3 and with the relative average of the WSS for DA varying from 50 to 34% for case 1, from 71 to 48% for case 2 and from 30 to 15% for case 3. To model the hemodynamics in intracranial aneurysm accurately, it is necessary to not only obtain high-resolution geometric information but also to use accurate flow information of the parent artery proximal to the intracranial aneurysm.

Using mainly commercially available computer workstations and software, these studies demonstrate the potential of CFD to study hemodynamics in cerebral aneurysms and their adjacent vascular segments. The precise role of flow velocity, dynamic pressure, static pressure and wall shear stresses and the mechanisms for aneurysm creation, aneurysm rupture or aneurysm re-growth, recanalization or re-bleeding after endovascular treatment, however, are still poorly understood. CFD methods could be used in the future to address these questions and to investigate patient-specific computational models in time frames suitable to impact therapeutic decision making (Fig. 6).

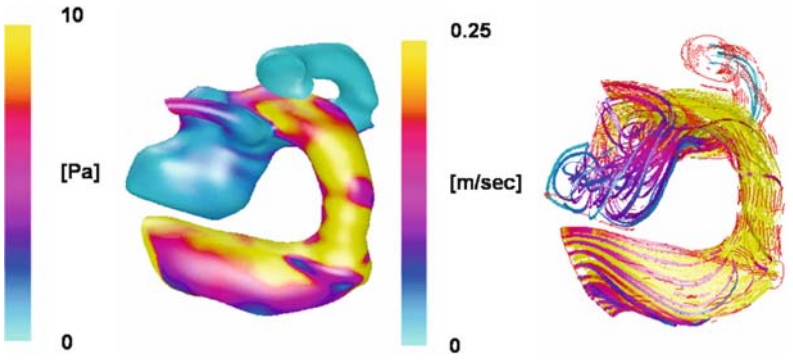


Fig. 6 WSS magnitude (on *left*) and pathlines (on *right*) of flow for an inflow velocity of 0.25 m/s obtained with the computational mesh shown in Fig. 5 (ANSYS Fluent, laminar unsteady solver)

Multi-modality Imaging and Modeling of Cerebral Aneurysms Hemodynamics

While earlier studies focused on the aneurysmal geometry only and modeled the inflow into the aneurysm by idealized waveforms, more recent studies integrated 3D DSA image data (for geometry) with 2D pcMRA data (for flow information) in a multi-modality approach.

The dependence of the blood flow pattern and the WSS magnitude on the inflow boundary conditions in CFD simulations was assessed through modeling an aneurysm of the anterior commuting artery (AComA) with inflow measured with phase contrast magnetic resonance (pcMRI) [63]. Variations in the intra-aneurismal blood flow patterns relative to patterns measured with pcMRI were investigated by varying the inflow conditions relative to each other in additional CFD simulations. Inflow volumetric waveforms into an AComA aneurysm were measured with phase contrast magnetic resonance on artery segments of the left and right anterior cerebral arteries proximal to the aneurysm and a transient CFD simulation was performed where the inflow was modeled from these measurements. To determine the effect of inflow boundary conditions deviating from the ones measured with pcMRI, the flow rate of the RACA was varied from 60 to 120% in five additional unsteady CFD simulations while adjusting the volumetric inflow in the LACA accordingly (from 168 to 66%) to keep total inflow constant. Good agreement was found between the intra-aneurismal velocity patterns simulated with CFD and the ones measured with pcMRI. Strong deviations in these patterns were noted when varying the inflow boundary conditions relative to each other in separate transient CFD simulations. Results indicate that the WSS magnitude in AComA aneurysms is sensitive to the exact inflow conditions and that pcMRI – by measuring these inflow conditions and the velocity patterns inside the aneurysms – may provide information to perform a more realistic simulation of the hemodynamics in this kind of cerebral aneurysm.

In three intracranial aneurysms, pcMRA in combination with the NOVA system was employed to visualize, and to quantify intra-aneurysm flow patterns in two approximately perpendicular scan planes inside each aneurysm [64]. The measured flow patterns were then compared to results obtained with CFD methods. The flow patterns obtained with pcMRI and CFD were found to be in good qualitative and quantitative agreement encouraging the use of CFD in for modeling of intra-aneurysmal hemodynamics. Differences between simulation and measurement are to be expected: pcMRA methods can only provide a partial-volume averaged image of the flow (slice thickness 5 mm), and CFD can only provide an approximation based on the chosen boundary conditions. Nevertheless, the main features of the flow (as calculated with CFD methods), especially of blood with opposite velocity directions in the scan planes were well reproduced in the simulations. pcMRA combined with NOVA may be able to provide a fast, non-invasive method to help classify intracranial aneurysms based on their flow patterns.

Flow patterns in an intracranial bifurcation aneurysm were calculated with CFD before and after embolization [65]. Simulated flow patterns were validated by comparison with 2D DSA time-resolved projection images acquired with a high

temporal resolution (seven images per second or higher). Excellent qualitative agreement between simulated CFD path lines and visualized flow patterns were found. Embolization created significant changes in the flow patterns, the wall shear stress, and dynamic pressure distributions. The results may have potential clinical applications in the optimization of endovascular treatment.

Outlook

The recent advances in computational techniques now enable true patient-specific modeling of the hemodynamics in cerebral aneurysms. CFD techniques were successfully combined to yield new insights into aneurysmal flow patterns and wall forces created by the flowing blood. Most notably, WSS was identified as a possible indicator that might be related to aneurysmal rupture risk. Improvements in imaging techniques will lead to a refinement of the simulation techniques, for example, cine magnetic resonance techniques now are capable of providing images in 100 ms or less, so that real-time information of aneurysm motion will be accessible. This information can then be used to derive the boundary conditions for fluid structure interactions [66]. Novel 4D pcMRA methods have been demonstrated to yield time-resolved 3D profiles of hemodynamics in various vascular beds [67, 68]. This method can be utilized to verify computational results. Surgical techniques will directly benefit from these advances. The interventionalist will obtain additional information through these simulations not accessible otherwise and most likely this information will lead to an improvement of patient care.

References

1. Sekhar LN, Heros RC (1981) Origin, growth, and rupture of saccular aneurysms: a review. Neurosurgery 8(2):248–260
2. van Gijn JRG (2001) Subarachnoid haemorrhage: diagnosis, causes and management. Brain 1(24):249–278
3. Deruty R, Pelissou-Guyotat I, Mottolese C, Amat D (1996) Management of unruptured cerebral aneurysms. Neurol Res 18(1):39–44
4. Kassell NF, Torner JC, Haley EC Jr, Jane JA, Adams HP, Kongable GL (1990) The international cooperative study on the timing of aneurysm surgery. Part 1: overall management results. J Neurosurg 73(1):18–36
5. Rasmussen PA, Mayberg MR (2004) Defining the natural history of unruptured aneurysms. Stroke 35(1):232–233
6. Mayer SA, Kreiter KT, Copeland D, Bernardini GL, Bates JE, Peery S, Claassen J, Du YE, Connolly ES Jr (2002) Global and domain-specific cognitive impairment and outcome after subarachnoid hemorrhage. Neurology 59(11):1750–1758
7. Hackett ML, Anderson CS (2000) Health outcomes 1 year after subarachnoid hemorrhage: an international population-based study. The Australian Cooperative Research on Subarachnoid Hemorrhage Study Group. Neurology 55(5):658–662

8. Ropper AH, Zervas NT (1984) Outcome 1 year after SAH from cerebral aneurysm. Management morbidity, mortality, and functional status in 112 consecutive good-risk patients. J Neurosurg 60(5):909–915
9. Broderick JP, Broat TG, Duldner JE, Tomsick T, Leach A (1994) Initial and recurrent bleeding are the major causes of death following subarachnoid hemorrhage. Stroke 25:1342–1347
10. Kassell NLG (2001) Unruptured intracranial aneurysms: in search of the best management strategy. Stroke 32:603–605
11. Leblanc R, Worsley KJ (1995) Surgery of unruptured, asymptomatic aneurysms: a decision analysis. Can J Neurol Sci 22(1):30–35
12. Chang HS, Kirino T (1995) Quantification of operative benefit for unruptured cerebral aneurysms: a theoretical approach. J Neurosurg 83(3):413–420
13. Raaymakers TW, Rinkel GJ, Limburg M, Algra A (1998) Mortality and morbidity of surgery for unruptured intracranial aneurysms: a meta-analysis. Stroke 29:1531–1538
14. King JT Jr, Berlin JA, Flamm ES (1994) Morbidity and mortality from elective surgery for asymptomatic, unruptured, intracranial aneurysms: a meta-analysis. J Neurosurg 81(6):837–842
15. van Crevel H, Habbema JD, Braakman R (1986) Decision analysis of the management of incidental saccular aneurysm. Neurology 36:1335–1339
16. Guglielmi G, Vinuela F, Dion J, Duckwiler G (1991) Electrothrombosis of saccular aneurysms via endovascular approach. Part 2: preliminary clinical experience. J Neurosurg 75(1):8–14
17. Raymond J, Roy D, Bojanowski M, Moumdjian R, L'Esperance G (1997) Endovascular treatment of acutely ruptured and unruptured aneurysms of the basilar bifurcation. J Neurosurg 86(2):211–219
18. Raymond J, Roy D (1997) Safety and efficacy of endovascular treatment of acutely ruptured aneurysms. Neurosurgery 41(6):1235–1245; discussion 1245–1236
19. Roy D, Raymond J, Bouthillier A, Bojanowski MW, Moumdjian R, L'Esperance G (1997) Endovascular treatment of ophthalmic segment aneurysms with Guglielmi detachable coils. AJNR Am J Neuroradiol 18(7):1207–1215
20. Willinsky RA (1999) Detachable coils to treat intracranial aneurysms. CMAJ 161(9):1136
21. Nepper-Rasmussen HJ, Andersen PB, Edal AL (1999) [Subarachnoid hemorrhage and intracranial aneurysms]. Ugeskr Laeger 161(45):6211
22. van Rooj WJ, Sluzewski M, Wijnalda D, Verhagen I, Schellens RL, op de Coul AA (1996) Intravascular treatment of inoperable cerebral aneurysm using Guglielmil's spirals; initial results in the Netherlands. Ned Tijdschr Geneeskd 140:491–495
23. Molyneux A, Kerr R (1999) International subarachnoid aneurysm trial. J Neurosurg 91(2): 352–353
24. Murayama Y, Vinuela F, Duckwiler GR, Gobin YP, Guglielmi G (1999) Embolization of incidental cerebral aneurysm by using the Guglielmi detachable coil system. J Neurosurg 90:207–214
25. Roy D, Milot G, Raymond J (2001) Endovascular treatment of unruptured aneurysms. Stroke 32(9):1998–2004
26. Jou LD, Mohamed A, Lee DH, Mawad ME (2007) 3D rotational digital subtraction angiography may underestimate intracranial aneurysms: findings from two basilar aneurysms. AJNR Am J Neuroradiol 28(9):1690–1692
27. Broderick JP, Viscoli CM, Brott T, Kernan WN, Brass LM, Feldmann E, Morgenstern LB, Wilterdink JL, Horwitz RI (2003) Major risk factors for aneurysmal subarachnoid hemorrhage in the young are modifiable. Stroke 34(6):1375–1381
28. Suarez JI, Tarr RW, Selman WR (2006) Aneurysmal subarachnoid hemorrhage. N Engl J Med 354(4):387–396
29. Weir B (2002) Unruptured intracranial aneurysms: a review. J Neurosurg 96(1):3–42
30. German W, Black P (1954) Intra-aneurysmal hemodynamics: turbulence. Trans Am Neurol Assoc 79:163–165
31. Karmonik C, Arat A, Benndorf G, Akpek S, Klucznik R, Mawad ME, Strother CM (2004) A technique for improved quantitative characterization of intracranial aneurysms. AJNR Am J Neuroradiol 25(7):1158–1161

32. Ujiie H, Tachibana H, Hiramatsu O, Hazel AL, Matsumoto T, Ogasawara Y, Nakajima H, Hori T, Takakura K, Kajiya F (1999) Effects of size and shape (aspect ratio) on the hemodynamics of saccular aneurysms: a possible index for surgical treatment of intracranial aneurysms. Neurosurgery 45(1):119–129; discussion 129–130
33. Ujiie H, Tamano Y, Sasaki K, Hori T (2001) Is the aspect ratio a reliable index for predicting the rupture of a saccular aneurysm? Neurosurgery 48(3):495–502; discussion 502–493
34. Black SP, Leo HL, Carson WL (1988) Recording and measuring the interior features of intracranial aneurysms removed at autopsy: method and initial findings. Neurosurgery 22(1 Pt 1):40–44
35. Ma B, Harbaugh RE, Raghavan ML (2004) Three-dimensional geometrical characterization of cerebral aneurysms. Ann Biomed Eng 32(2):264–273
36. Hoi Y, Meng H, Woodward SH, Bendok BR, Hanel RA, Guterman LR, Hopkins LN (2004) Effects of arterial geometry on aneurysm growth: three-dimensional computational fluid dynamics study. J Neurosurg 101(4):676–681
37. Castro MA, Putman CM, Cebral JR (2006) Patient-specific computational fluid dynamics modeling of anterior communicating artery aneurysms: a study of the sensitivity of intra-aneurysmal flow patterns to flow conditions in the carotid arteries. AJNR Am J Neuroradiol 27(10): 2061–2068
38. Cebral JR, Castro MA, Burgess JE, Pergolizzi RS, Sheridan MJ, Putman CM (2005) Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. AJNR Am J Neuroradiol 26(10):2550–2559
39. Dempere-Marco L, Oubel E, Castro M, Putman C, Frangi A, Cebral J (2006) CFD analysis incorporating the influence of wall motion: application to intracranial aneurysms. Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv 9(Pt 2): 438–445
40. Di Martino ES, Guadagni G, Fumero A, Ballerini G, Spirito R, Biglioli P, Redaelli A (2001) Fluid-structure interaction within realistic three-dimensional models of the aneurysmatic aorta as a guidance to assess the risk of rupture of the aneurysm. Med Eng Phys 23(9):647–655
41. Hoi Y, Woodward SH, Kim M, Taulbee DB, Meng H (2006) Validation of CFD simulations of cerebral aneurysms with implication of geometric variations. J Biomech Eng 128(6):844–851
42. Jou LD, Quick CM, Young WL, Lawton MT, Higashida R, Martin A, Saloner D (2003) Computational approach to quantifying hemodynamic forces in giant cerebral aneurysms. AJNR Am J Neuroradiol 24(9):1804–1810
43. Karmonik C, Benndorf G, Klucznik R, Haykal H, Strother CM (2006) Wall shear stress variations in basilar tip aneurysms investigated with computational fluid dynamics. Conf Proc IEEE Eng Med Biol Soc 1:3214–3217
44. Karmonik C, Klucznik R, Benndorf G (2008) Comparison of velocity patterns in an AComA aneurysm measured with 2D phase contrast MRI and simulated with CFD. Technol Health Care 16(2):119–128
45. Karmonik C, Klucznik R, Benndorf G (2008) Blood flow in cerebral aneurysms: comparison of phase contrast magnetic resonance and computational fluid dynamics – preliminary experience. Rofo 180(3):209–215
46. Mantha A, Karmonik C, Benndorf G, Strother C, Metcalfe R (2006) Hemodynamics in a cerebral artery before and after the formation of an aneurysm. AJNR Am J Neuroradiol 27(5):1113–1118
47. Steinman DA, Milner JS, Norley CJ, Lownie SP, Holdsworth DW (2003) Image-based computational simulation of flow dynamics in a giant intracranial aneurysm. AJNR Am J Neuroradiol 24(4):559–566
48. Benndorf G, Singel S, Proest G, Lanksch W, Felix R (1997) The Doppler guide wire: clinical applications in neuroendovascular treatment. Neuroradiology 39(4):286–291
49. Pelc NJ, Herfkens RJ, Shimakawa A, Enzmann DR (1991) Phase contrast cine magnetic resonance imaging. Magn Reson Q 7(4):229–254
50. Pipe JG (2003) A simple measure of flow disorder and wall shear stress in phase contrast MRI. Magn Reson Med 49(3):543–550

51. Bluestein D, Niu L, Schoephoerster RT, Dewanjee MK (1996) Steady flow in an aneurysm model: correlation between fluid dynamics and blood platelet deposition. J Biomech Eng 118(3):280–286
52. Tateshima S, Murayama Y, Villablanca JP, Morino T, Takahashi H, Yamauchi T, Tanishita K, Vinuela F (2001) Intraaneurysmal flow dynamics study featuring an acrylic aneurysm model manufactured using a computerized tomography angiogram as a mold. J Neurosurg 95(6): 1020–1027
53. Tateshima S, Tanishita K, Omura H, Villablanca JP, Vinuela F (2007) Intra-aneurysmal hemodynamics during the growth of an unruptured aneurysm: in vitro study using longitudinal CT angiogram database. AJNR Am J Neuroradiol 28(4):622–627
54. Baumgartner RW (1999) Transcranial color-coded duplex sonography. J Neurol 246(8): 637–647
55. Gailloud P, Khan HG, Albayram S, Martin JB, Rufenacht DA, Murphy KJ (2002) Pooling of echographic contrast agents during transcranial Doppler sonography: a sign in favor of slow-flowing giant saccular aneurysms. Neuroradiology 44(1):21–24
56. Shojima M, Oshima M, Takagi K, Torii R, Hayakawa M, Katada K, Morita A, Kirino T (2004) Magnitude and role of wall shear stress on cerebral aneurysm: computational fluid dynamic study of 20 middle cerebral artery aneurysms. Stroke 35(11):2500–2505
57. Malek AM, Izumo S (1995) Control of endothelial cell gene expression by flow. J Biomech 28(12):1515–1528
58. Jou LD, Wong G, Dispensa B, Lawton MT, Higashida RT, Young WL, Saloner D (2005) Correlation between lumenal geometry changes and hemodynamics in fusiform intracranial aneurysms. AJNR Am J Neuroradiol 26(9):2357–2363
59. Acevedo-Bolton G, Jou LD, Dispensa BP, Lawton MT, Higashida RT, Martin AJ, Young WL, Saloner D (2006) Estimating the hemodynamic impact of interventional treatments of aneurysms: numerical simulation with experimental validation: technical case report. Neurosurgery 59(2):E429–E430; author reply E429–E430
60. Cebral JR, Pergolizzi RS Jr, Putman CM (2007) Computational fluid dynamics modeling of intracranial aneurysms: qualitative comparison with cerebral angiography. Acad Radiol 14(7):804–813
61. Karmonik C, Benndorf G, Haykal H, Klucznik R (2007) Comparison of blood inflow patterns into AComA aneurysms determined with 3D DSA and simulated with CFD. Chicago, IL
62. Karmonik C, Benndorf G, Klucznik R, Haykal H, Strother CM (2006) Wall shear stress variations in basilar tip aneurysms investigated with computational fluid dynamics. New York, NY
63. Karmonik C, Yen C, Grossman RG, Klucznik R, Benndorf G (2009) Intra-aneurysmal flow patterns and wall shear stresses calculated with computational flow dynamics in an anterior communicating artery aneurysm depend on knowledge of patient-specific inflow rates. Acta Neurochir (Wien) 151(5):479–485
64. Karmonik C, Benndorf G, Klucznik RHH, Strother C (2006) Comparison of cerebral intra-aneurysm flow characteristics determined by phase contrast magnetic resonance imaging and computational fluid dynamics simulations. New York, NY. ISMRM
65. Karmonik C, Morsi H, Mawad M (2005) Flow patterns in intracranial aneurysms before and after treatment – simulated with computational fluid dynamics. p 144
66. Tezduyar TE, Sathe S, Cragin T, Nanna B, Conklin BS, Pausewang J, Schwaab M (2007) Modelling of fluid-structure interactions with the space-time finite elements: arterial fluid mechanics. Int J Numer Methods Fluids 54(6–8):901–922
67. Frydrychowicz A, Berger A, Russe MF, Stalder AF, Harloff A, Dittrich S, Hennig J, Langer M, Markl M (2008) Time-resolved magnetic resonance angiography and flow-sensitive 4-dimensional magnetic resonance imaging at 3 Tesla for blood flow and wall shear stress analysis. J Thorac Cardiovasc Surg 136(2):400–407
68. Harloff A, Albrecht F, Spreer J, Stalder AF, Bock J, Frydrychowicz A, Schollhorn J, Hetzel A, Schumacher M, Hennig J, Markl M (2009) 3D blood flow characteristics in the carotid artery bifurcation assessed by flow-sensitive 4D MRI at 3T. Magn Reson Med 61(1):65–74

A Computational Framework for Breast Surgery: Application to Breast Conserving Therapy

David Thanoon, Marc Garbey, Nam-Ho Kim, and Barbara Bass

Problem and Motivation

Introduction

If a surgical intervention is needed, early stage breast cancer may lead to three basic surgery choices: breast-sparing surgery followed by radiation therapy, mastectomy, mastectomy with breast reconstruction surgery. Breast-sparing surgery (breast conservation therapy (BCT)) removes the breast tumor and a margin of surrounding normal tissues. It is also known by other names: lumpectomy, partial mastectomy, segmental mastectomy, and quadrantectomy. Radiation therapy follows lumpectomy to eliminate any microscopic cancer cells in the remaining breast tissue. The purpose of BCT is to give women the same cure rate they would have if they were treated with a mastectomy but to leave the breast intact, with an appearance and texture as close as possible to what they had before treatment. Trials for breast conservative therapy with patients affected by breast cancer (I and II) have demonstrated conclusively that BCT produces disease control and survival rates at least equivalent to those of mastectomy, and possibly better in the long run for patients with stage II [1]. BCT is combined to radiation therapy. While BCT removes the tissue that contain the tumor with a negative margin, radiotherapy insure that residual microscopic disease are controlled. Contra-indications to BCT are generally for patients with high probability of recurrence, high probability of normal tissue damage from irradiation. While cosmesis after BCT might be generally satisfactory, the

D. Thanoon (✉) and M. Garbey
Department of Computer Science, University of Houston, 501 Phillip Hoffman Hall, Houston, TX 77204, USA
e-mail: dthanoon@cs.uh.edu

N.-H. Kim
Department of Mechanic and Aerospace Engineering, University of Florida, Gainesville, FL, USA

B. Bass
Department of Surgery at The Methodist Hospital, Houston TX, and Weill Medical College of Cornell University, New York, NY, USA

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_15,

quality of the result is very sensitive to the location and extent of the tumor. Further, the breast is a very deformable structure with a complicated anatomy that is patient specific. The mechanical properties of glandular, fatty and cancerous tissue are quite different, and vary from one patient to another. The Cooper's ligament also plays a key role in the outcome. Strong asymmetry in the location or large tumor size are prone to anesthetic BCT result. Surgical results are also depending on the time scale. Beside the short time scale modeling that might be caught by the mechanical model, one can expect that inflammation, postsurgical radiotherapy, and healing dynamic can change significantly in the long time scale of the cosmetic outcome. In other words biology plays a role as well.

The same way that it takes a multidisciplinary team, involving a surgeon, a radiologist, an oncologist, etc. to manage breast cancer, it takes a multidisciplinary model involving soft tissue mechanic, medical imaging, and biology modeling to predict the outcome of BCT. The objectives of this study are to construct a modular computational model that can allow the prediction of the breast surface contour, after partial mastectomy and potentially identify targets for intervention to improve cosmetic results. Key questions are

- A priori model of the breast deformation in order to be able to define the pattern of cosmetic defects for women undergoing BCT with numerical simulation.
- To determine if patterns of deformity can be predicted based on preoperative imaging and surgical data points.

The philosophy of this preliminary study is to develop a computational framework that can feed on a cost-effective clinical protocol and help significantly surgery planning with BCT. We do not look for a perfect model but rather a useful model.

First, we are going to review the individual models that are available in the literature.

Model and Related Work

There has been a large stream of work on patient specific bio-mechanical models of the breast used to predict deformation [2]. The goal of such studies is to compute the change in shape as external mechanical forces are varied. The standard application is co-registration of internal structure of interest such as tumors. We are going to see that this soft tissue mechanical problem by itself is fairly difficult to solve. While the surgical intervention is typically with the patient lying on his back, presurgery planning relies on medical imaging with different positions. The breast is a highly deformable organ fairly sensitive to gravity load. For MRI, the patient lies prone and the breast is allowed to drop under gravity. The prone position is not particularly inherent to the MRI procedure, but rather to the available existing equipment. It is unfortunate that these MRI images cannot be directly used to surgery but should rather be transformed to accurately track the tumor location. For X-ray, the subject stands and the breast has to be pressed between two plates. Under such conditions, it is critical to compute an accurate registration of the tumor that can localize the

tumor in the absence of external pressure. Registration should be much improved if one uses a bio-mechanical model to predict the tumor location as a function of the gravity load. The problem of an accurate registration of the tumor location is also seen in biopsy because the needle is not a very sharp object. A straight trajectory may result in fact into a trajectory that changes direction at the interface between different layers of tissues. Overall, the biopsy may not collect the targeted tissue sample. There is a large number of papers on the prediction of mechanical deformation of the breast that started with the extensive use of finite element models (FEM) [3–5]; two papers of particular interest are the Medical Image Analysis article of Azar et al. [6] that concentrates on the biopsy application and emphasizes the link between the model and image analysis, and the article of Rajagopal et al. [7] that discusses the problem of the reference state. Both papers have pretty extensive bibliographies. In Azar et al. work, a custom written program for the image segmentation of the 3D breast is linked to the ABAQUS model. Image segmentation of the breast should at least take into account the two predominant types of tissue that are fat and normal glandular tissue. Image segmentation of the tumor should be added for BCT.

There is a fair amount of uncertainty in material properties of the soft tissue that composes the breast. Further, such properties are very much patient specific. According to Krouskop et al. [8], fat tissue and glandular tissue have similar elastic modulus at low strain levels, while glandular tissue elastic modulus increases by one order of magnitude at high strain levels. Further, fibrous tissue have an elastic modulus one to two order of magnitude higher than fat tissue. In [6], the breast fatty tissue model takes also into account the effect of fat compartmentalization due to Cooper's ligament. Finally, tumor tissues are much stiffer than surrounding tissue as measured for example in [9].

Actually, skin mechanical properties might also be taken into account as a separate and modeled possibly as a two-dimensional deformable shell that bounds the 3D model. Further, it might be difficult to predict the right boundary conditions that describe the attachment of the breast to the muscle of the thorax. For all these reasons, while the validation of the FEM with a phantom study using a deformable silicon gel is kind of reassuring, the FEM still needs to be validated with patient as in [6, 10]. Most previous studies of breast modeling deformations have been conducted for large deformation due to mammographic compression or for biopsy. Most of these studies were conducted using Mooney–Rivlin Hyperelastic models under large deformation. Still, small deformation theory has been used thanks to either a continuation method or time stepping but requires then to control the accumulation of errors from each step. In many of these studies, the computation of the mechanical model was verified by comparison with experimental data on gel phantoms that have very similar properties to breast tissue. Comparison to clinical data is however sparse and/or involves very few human subjects. One additional difficult problem is that one cannot recover directly the unloaded shape of the breast. Rajagopal et al. [7] have shown that the accuracy of shape prediction is much improved by recovering a stress free unloaded position virtually. In this work, Rajagopal et al. propose a numerical algorithm to solve that inverse problem, and use a special experimental set up to validate that study with human subjects. One still needs to assume that the unloaded shape is free of stress, which is fairly unclear in particular for breast

cancer. Though most of these studies reported here considered the breast tissue as incompressible isotropic homogeneous, no study, as far as we know, was focussed on BCT.

Realistic tumor dynamic modeling is very challenging as well. For BCT, one is interested on tissue margin, i.e., how much of the tissue BCT should take away, and possibly how one can make sure that chemotherapy and/or radiotherapy can control what might be left of cancer cells in the breast. There is a vast literature on model and simulation of tumor growth. Some with multiscale PDE model [25]. Cellular automata (CA) [11] and/or agent base model (ABM) [12] might be used as well to simulate the dynamic of the population of cells involved in cancer [13], in particular in situation with sparse density of tumor cells. Both CA and ABM techniques belong to the bottom–up approach where individual cells behavior and interaction with their environment are explicitly modeled. One can observe a posteriori in the simulation the emerging dynamic of the complex system. Implementation of these methods has been reviewed recently in [14]. These methods can be applied to tumor dynamic as well as modeling inflammation and tissue healing [15–19]. The limit of such model is the inherent complexity of the biological environment and the difficulty to capture adaptation and mutation of individual cells.

However, the future is to design multiscale hybrid model that can couple the mechanical description of tissue with the long time scale dynamic of the biological tissue that is essential for recovery [20]. The challenge, indeed, is to have patient specific model that can be used in clinical conditions and that can feed on existing data acquisition processes such as tissue elastography for tissue mechanical properties and biopsy for cell population.

Now, we are going to discuss various computational methods to simulate the mechanical deformation of the breast. In this chapter, we will report exclusively on artificial benchmark problems to develop the computational methods. Image-based simulation with patient data will be reported elsewhere.

Model of Mechanical Deformation of the Breast

Typically, a mechanical model requires the description of the material properties of the breast along with anatomic data, an initial state and some boundary conditions.

The model can be either static, i.e., describe an equilibrium, or dynamic, i.e., describes the time dependent deformation. Further, the model can incorporate the material properties of the tissue with a wide range of level of details.

We are going to review our experiment with several models with increasing level of complexity. We will assume that we start from a known shape at zero gravity conditions and with zero internal stress. The solution of our benchmark problem is the displacement and stress distribution for the equilibrium position at standard gravity. All simulations here are implemented with the commercial software COMSOL MULTIPHYSICS™ [21]. COMSOL is a software based on finite element approximation that allows fast prototyping.

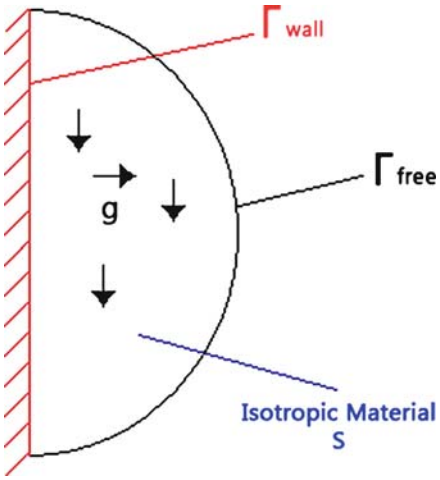


Fig. 1 Geometry of the benchmark problem

For simplicity, our benchmark problem is a semisphere attached to the wall as in Fig. 1. This simplified benchmark problem will be used to compare linear and nonlinear models, and will be enough to show some limitation of the simulation. We start with the standard linear elasticity model.

Linear Elasticity

Let us start with the small deformation theory that is implicitly assumed in the linear elasticity model which will help is understand further refinement in the modeling: The unknown of the mechanical problem is the field:

$$\left(\overrightarrow{u}\left(\overrightarrow{x}\right), \overline{\overline{\sigma}}\left(\overrightarrow{x}\right)\right), \quad \forall \overrightarrow{x} \in S \tag{1}$$

in the semisphere S with $\overrightarrow{u}\left(\overrightarrow{x}\right)$ representing the displacement, $\overline{\overline{\sigma}}\left(\overrightarrow{x}\right)$ representing the engineering stress.

It is possible to completely describe the strain condition at a point with the deformation components u_i for$i = 1, 2, 3$ and their derivatives. The strain components are given from the deformation as follows:

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{2}$$

Also, the stress–strain relationship for a linear isotropic material reads:

$$\overline{\overline{\sigma}} = \lambda\left(tr\overline{\overline{\varepsilon}}\right)\overline{\overline{I}}_3 + 2\mu\,\overline{\overline{\varepsilon}}, \quad \forall \overrightarrow{x} \in S \tag{3}$$

with λ and μ the elasticity coefficient of Lame linked to E, the modulus of elasticity (known as *Young's modulus*), and ν *Poisson's ratio* by

$$\lambda = E\frac{\nu}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{2(1+\nu)}. \tag{4}$$

Under static conditions, the equilibrium equations write:

$$\operatorname{div}\overline{\overline{\sigma}} + \rho\overrightarrow{g} = \overrightarrow{0}, \tag{5}$$

where $\overrightarrow{g}$ denotes the body forces per mass.

The boundary conditions are the following:

For the embedment: $u_i\left(\overrightarrow{x}\right) = 0, \quad \forall\overrightarrow{x} \in \Gamma_{\text{wall}}$

For the free surface: $t_i\left(\overrightarrow{x}\right) = \sigma_{ij}\left(\overrightarrow{x}\right)n_j\left(\overrightarrow{x}\right) = 0, \quad \forall\overrightarrow{x} \in \Gamma_{\text{free}}$

With $\Gamma = \Gamma_{\text{wall}} \cup \Gamma_{\text{free}}, \quad \Gamma_{\text{wall}} \cap \Gamma_{\text{free}} = \oslash$, where the $n_j(\overrightarrow{x}), (j = 1, 2, 3)$ are the components of the outward unit normal vector $\overrightarrow{n}(\overrightarrow{x})$.

Instead of directly using this formulation, to extend the small deformation theory to the large deformation situation, we use time stepping with the adaptive time step provided by COMSOL, and increase the gravity progressively from 0 at initial stage to g in a physical time of 60 s.

We suppose that the tissue is a linear isotropic material with *Young's Modulus* $E = 2{,}520$ Pa and a *Poisson's ratio* $\nu = 0,4$ [22]. The simulation is completed with 17,000 quadratic Lagrange elements. The static solution is shown in Fig. 2. Figure 2 shows the von Mises stress: a singularity appears near the corner of the breast. This singularity appears to be an artifact because of the nonrealistic boundary condition imposed at the "wall." This problem should be avoided by choosing a new geometry model that includes the body core. We provide more realistic boundary conditions next by adding a wall support as in Fig. 3. Regardless of this singularity, the breast static state has some analogy with good shape of a healthy women's breast.

The next step was to simulate the breast state after a surgical intervention of the breast with the removal of cancerous tissue. This surgical intervention was modeled with the introduction of a small sphere S_{removal} with a radius $R_{\text{interior}} = 3$ cm inside the semisphere – see Fig. 3 (material 2). We suppose that after the surgical intervention, this empty part will be first filled up with a liquid.

Model I: We approach the liquid properties material by choosing a water density $\rho = 1{,}000\,\text{kg m}^{-3}$ a small *Young's Modulus* $E = 300$ Pa and a *Poisson's ratio* $\nu = 0{,}495$ since only the mass contribution of the fluid during the analysis is of interest. Simulation results are shown in Fig. 4.

This simulation still exhibits a singularity near the wall due to the change in material properties at the interface between materials 2 and 3. We observe the impact of tissue removal on the overall shape of the breast. A bump appears which gives a nonconform shape to the breast. Our modeling of tissue removal is fairly naive and the modeling of water can be interpreted in a more physical way.

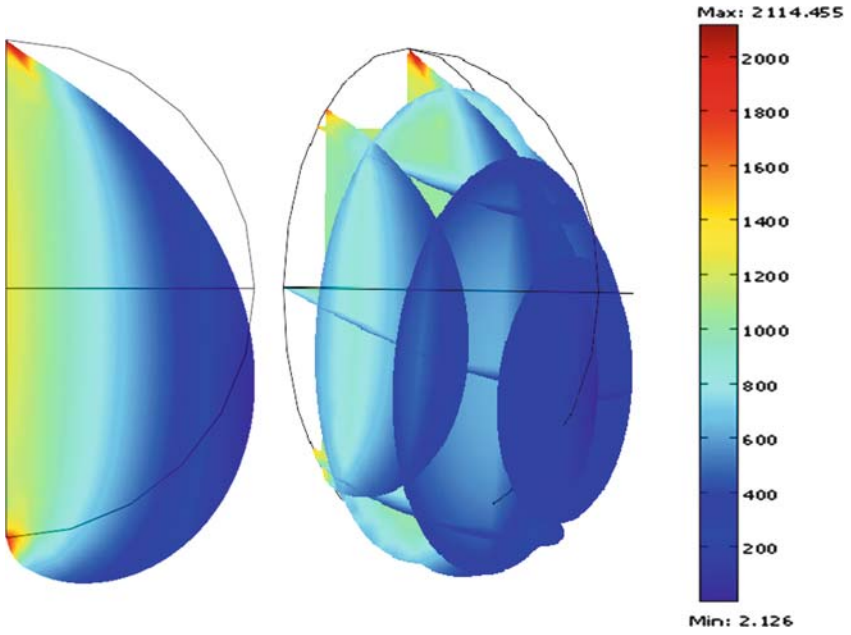


Fig. 2 Deformation of the single 3D isotropic breast model under the gravity, plotting the von Miss stress (Pa) ($E = 2{,}520\,\text{Pa}, \nu = 0,4$)

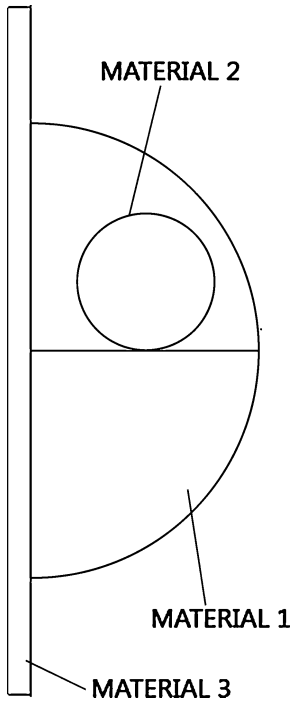


Fig. 3 Geometry of the postsurgery breast model

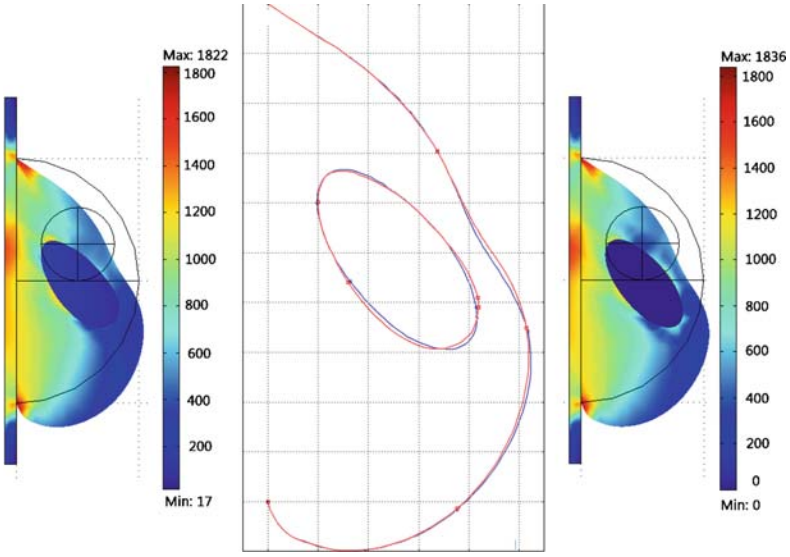


Fig. 4 Von miss stress and deformation of the breast model with tumor removal using: (*left*) water model I (min 17 Pa; max 1,822 Pa); (*right*) water model II (min 0 Pa; max 1,836 Pa); (*middle*) Comparison of breast and tumor removal deformation for water model I, and water model II gives very close by curves

Model II: To model water in COMSOL, we use an anisotropic material with a singular representation of the elasticity matrix. The stress–strain relationship for an anisotropic material is linked by the following relation $\sigma = D\ \epsilon$, where D is the 6×6 elasticity matrix and the stress and strain are both given in column vector form:

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{xz} \end{pmatrix} \text{ and } \begin{pmatrix} \epsilon_x \\ \epsilon_y \\ \epsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{xz} \end{pmatrix}$$

Water is supposed to be a perfect fluid, with no friction, indeed the only permissible load is the hydrostatic pressure, volumetric force. In order to avoid the friction force, the elasticity matrix is divided in four smaller 3×3 matrix where the three blocs that are in relation with the friction effect, τ_{**}, are equal to zero. Since only the hydrostatic pressure force occurs, it can be represented by the relation with the bulk modulus. Indeed, the bulk modulus of a fluid measures the fluid resistance to uniform compression. It is defined as the pressure increase needed to effect a given relative decrease in volume. The bulk modulus K can be formally defined by the equation: $K = -V\frac{\partial P}{\partial V}$ where P is the pressure, V is the volume, and $\frac{\partial P}{\partial V}$ denotes the partial derivative of pressure with respect to volume. The inverse of the bulk

modulus gives a substance compressibility. This relation can be arranged as the following $K\frac{\partial V}{V} = -\partial P$. By analogy on mechanic basic, we can state that $\sigma = -\partial P$ and that trace$(\epsilon) = (\epsilon_x + \epsilon_y + \epsilon_z) = \frac{\partial V}{V}$. This leads to an elasticity matrix with the following form:

$$D = \begin{pmatrix} K & K & K & 0 & 0 & 0 \\ K & K & K & 0 & 0 & 0 \\ K & K & K & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The result of this simulation is illustrated in Fig. 4, where it shows von Miss stress of the meridional cut of the breast. This simulation was done with the same elastic linear model than before and the value of the water bulk modulus $K = 2.2e9\,\text{Pa}$. In comparison with the first water model, we can see that in our second model the value of the stress inside the liquid is equal to zero, whereas in the first model the liquid has some stress. We can assume that our second model is more realistic and gives more physical result. However, we can see that in the two models the bump and the exterior shape are quite similar.

The material property chosen for the model so far was linear isotropic. This is not the true mechanical properties of soft tissue. Indeed, the stress–strain relationship of breast tissue is hyperelastic.

Toward a Nonlinear Modeling

The breast often deforms significantly, linear elasticity with the infinitesimal deformation formulation is not appropriate to formulate the breast model. As a result, we used a finite deformation formulation in conjunction with hyperelastic material.

The problem is to find the coordinates (**x**) of the deformed body, V, given the coordinates (**X**) of the undeformed body V_0. The deformation gradient tensor **F** provides the relationship to map between the undeformed states, and is defined as

$$\overline{\overline{F}} = \left\{ \frac{\partial x_i}{\partial X_M} \right\}. \tag{6}$$

The Green Lagrange strain tensor **E** is calculated using:

$$\overline{\overline{E}} = \frac{1}{2}\left(\overline{\overline{F}}^T \overline{\overline{F}} - \overline{\overline{I}} \right). \tag{7}$$

The aim is to find a solution vector $\overrightarrow{x}$ representing the degrees of freedom defining the deformed state, such that the principles of conservation of mass, linear momentum, and angular momentum are satisfied. The modeling framework has been expressed with respect to the reference configuration and thus the equations below

are written in terms of the undeformed state. The reason is that we do not know the deformed shape since we are dealing with large deformation.

When a body is in equilibrium, all of the forces must balance. This is achieved by satisfying the principle of conservation of linear momentum that can be written for a static problem as:

$$\frac{\partial}{\partial X_M}\left(S^{\mathrm{MN}}\frac{\partial x_j}{\partial X_N}\right)+\rho_0 f^j=0, \tag{8}$$

where S^{MN} are components of the second Piola–Kirchhoff stress tensor (force per unit area of the undeformed body) and f_j are the body forces in the reference frame (such as gravity in our problem).

We model the breast tissue as a hyperelastic material which is a material where the stresses are computed from a strain energy density function. It is assumed that the Second Piola–Kirchhoff stress $\overline{\overline{S}}$ is used, so that

$$\overline{\overline{S}}=\frac{\partial W_S}{\partial \overline{\overline{E}}}.$$

For an isotropic material, the strain energy density function, W_S, can only be a function of the strain invariants. In a total Lagrangian formulation, it is convenient to use the right Cauchy–Green tensor $\overline{\overline{C}}=\overline{\overline{F}}^T\overline{\overline{F}}$ for the representation of the strain. Hence, the invariants are as follows:

$$I_1=\mathrm{trace}(C)=C_{11}+C_{22}+C_{33}, \tag{9}$$

$$I_2=\frac{1}{2}(I_1^2-\mathrm{trace}(C^2)), \tag{10}$$

$$I_3=\det(C)=J^2, \tag{11}$$

where $J=\det(\mathbf{F})$. We used Mooney–Rivlin strain energy, with no thermal expansion, to model breast tissue, which is defined as

$$W_S=C_{10}(-3+J^{-\frac{2}{3}}I_1)+C_{01}(-3+J^{-\frac{4}{3}}I_2)+\frac{\kappa}{2}(J-1)^2, \tag{12}$$

where C_{10}, C_{01} and κ are the material constants of the breast tissue, these constant units are Pa. The value of the material constants C_{10} and C_{01} for the breast tissue are to be determined experimentally. κ is the bulk modulus and carries the volumetric part that represents the resistance to hydrostatic compression. Therefore, the higher κ is the more incompressible is the material.

The simulations are done for a static problem. We set the value for the material constants to be $C_{10}=3{,}740\,\mathrm{Pa}$ and $C_{01}=1{,}970\,\mathrm{Pa}$; these values correspond to a material that is a silicon gel [6]. The first computation was made with a 14,000 quadratic Lagrange elements. This computation appears then to diverge while we checked that this was not an issue with mesh refinement. An easy solution is to

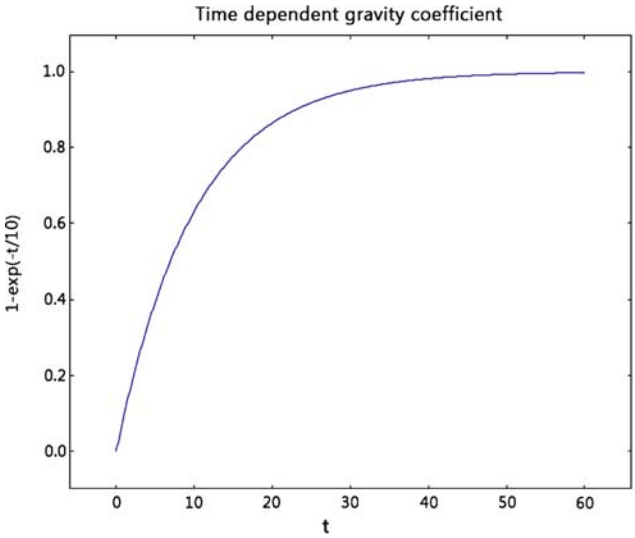


Fig. 5 Plot of the time dependent gravity load $\times 10\,\mathrm{m\,s^{-2}}$

compute the solution of this nonlinear problem as the end result of a transient problem with increasing gravity load as in Fig. 5. Alternatively, we can stay within the small deformation theory by artificially increasing the gravity field form zero to the standard value. This continuation method has two advantages: for each step, it keeps the strain below the small deformation criteria (strain under 4%) and it enables us to get a solution that converges for each step. Keeping small deformations between each step leads to fast computation. However, this method could lead to an accumulation of discretization errors. This method could be seen as a quasi-static method. In our simulation we have increased the gravity for each step by 0.3% of the real gravity value. This means that after 333 incremental steps we can reach the real gravity load. The result of this simulation is shown in Fig. 6. We can see no singularity at the corner and the von Miss stress is continuous. Nevertheless, the value of the stress is much higher than the linear elastic model. Although hyperelasticity and linear elastic model are not the same, there exists some connection. Indeed for low strain hypothesis (strain $< 5\%$) one can express the relationship between C_{01}, C_{10} and E. The material properties of silicon gel we use, did not allow us to link these two models. Also, more validation and verification against the experimental data need to be done to validate the model.

Influence of Poisson Ratio

One of the main difficulties with the breast mechanical model is the lack of initial conditions, i.e., the fact that no shape under zero load is available. A naive approach

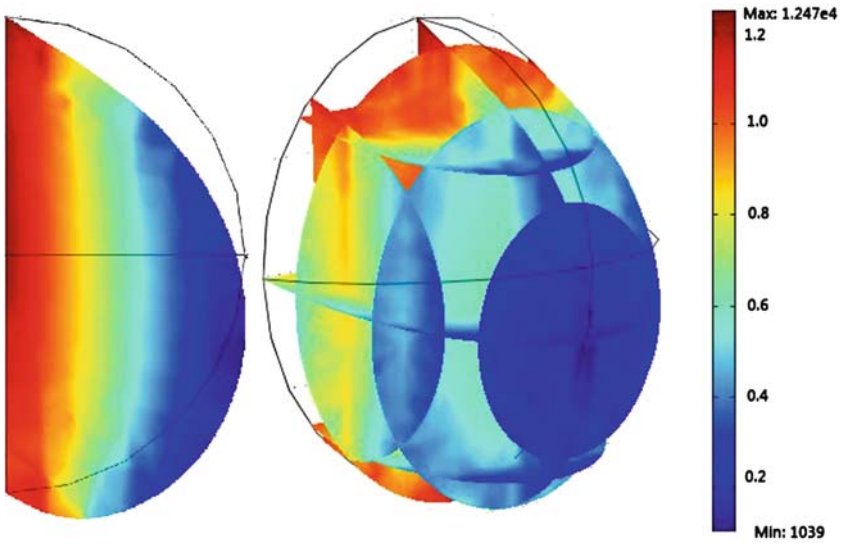


Fig. 6 Simulation of the deformation of the 3D breast model under the gravity using hyperelastic model and the "quasi-static" method, plotting the von Miss stress (max = 1.24e4 Pa; min = 1,039 Pa)

will be to start from the loaded configuration, reverse the gravity and expect that this will be the initial shape at zero gravity.

One of the reasons for which this idea does not work, beside the fact that the loaded configuration is not stress free, is that the standard linear elasticity model does not even conserve the volume. In the elastic linear equation, the poisson's ratio ν represents the incompressibility of a material. Indeed, when a sample of material is stretched in one direction, it tends to contract (or rarely, expand) in the other two directions. The Poisson's ratio, ν, is a measure of this tendency. The influence of the poisson's ratio can be shown by a two step method. First, we apply to our initial geometry a gravity load in a direction, e_z, and we obtain the geometry of the breast under a gravity load. The second step consists in extracting the deformed geometry under gravity and in applying a gravity load in the opposite direction, $-e_z$. After the first deformation, the volume changes, hence the integration of the volumetric forces for our deformed geometry differs from the original if we apply the same body force.

However, if our models were correct, we should obtain after the second step the exact same initial geometry. In order to see the influence of ν, we apply this two step procedure, for different value of ν. The Poisson ratio for materials stands between -1 and 0.5. The value of $\nu = 0.5$ representing the complete incompressibility cannot be reached according to (4). One compute the percentage of the relative error in L_2 norm for the meridional plan of the breast as a function of the second gravity load (in the inverse direction) for different values of ν. We observe that the closer ν is to 0.5 the smaller the error, but this error is no

less than 2% at minimum. A remedy to that lack of volume conservation would be to constraint the model with a Lagrange multiplier. Now, we are going to discuss an algorithm to solve the inverse problem that consists in the reconstitution of the initial unloaded shape from a loaded configuration.

Retrieving the Unloaded Shape

For simplicity, we restrict ourselves to a two-dimensional case.

Let us consider a curve $C_{\text{med}}(t) = (\theta_{\text{med}}(t), r_{\text{med}}(t))$ define for $t \in [0, \pi]$, that represents the 2D breast external envelop under gravity loading. This parametric representation is done in the polar coordinate system. Our goal is to find the curve $C_s(t) = (\theta_s(t), r_s(t))$ representing the breast model under zero gravity load. When we apply the gravity load to C_s, we should obtain C_{med}.

We consider a set of N_c points : $t_1 = 0 < t_1 < \cdots < t_{N_c} = \pi$ to support a B-spline representation c_s of our unknown contour curve C_s. Our optimization problem in this finite approximation space is summarized in Fig. 7.

The optimization algorithm is an iterative process that requires an initial guess C_0. We build the initial guess by applying the gravity load in the inverse

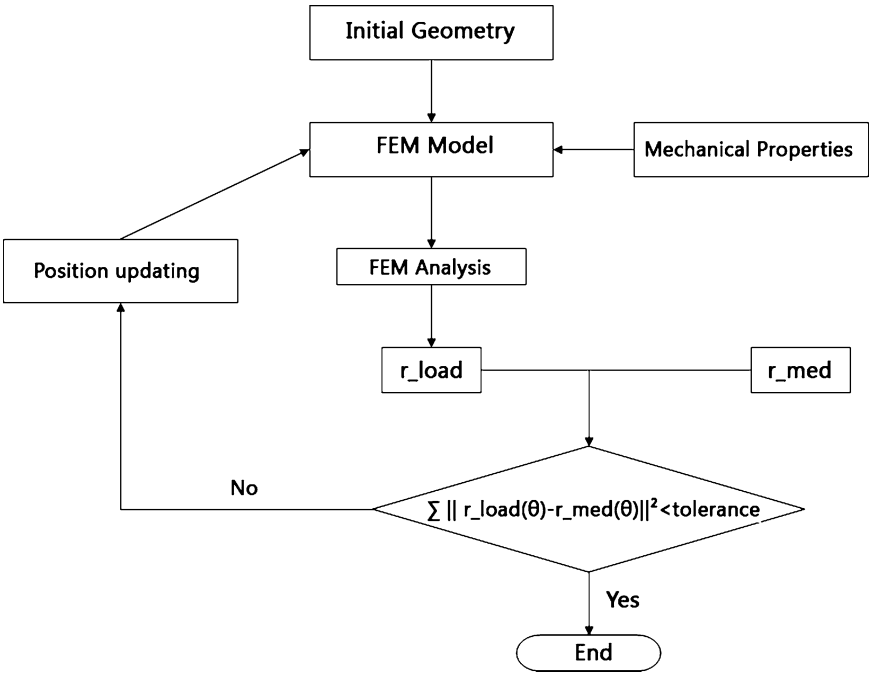


Fig. 7 Optimization algorithm flowchart used to determine the unloaded shape of a breast from a loaded shape

direction on C_{med}. We may apply a surface rescaling to impose the surface area to be preserved. After construction of our initial guess, we compute the loaded shape with one of the previous forward model and compare our controlling points to the medical shape, C_{med}. The controlling points position are updated with an optimization algorithm that minimizes the distance between the two curves. If the tolerance condition is met, the optimization process is terminated. This occurs when the sum of differences between the medical shape and the FEM shape is sufficiently small. Our first attempt to use a gradient method fails, because this method appears to be too sensitive to the noise of the forward FEM calculation. Second, we use a Nelder–Mead method [23] that turns out to be more stable and provided convergence. We found it to be more efficient to use some kind of a local relaxation technique that sweep the parameter space and solve the optimization problem of one component at a time. This simple algorithm is described as follows:

Until convergence criteria is met, repeat the sweeping algorithm:
for $i = 1 : N_c$

- *Optimize position of point* i *in* C_0
- *Set the new* C_0 *with the optimize position of point* i

end for

This procedure converges faster with the following regularization that consists to update the two neighbor points with a spline interpolation from the other fixed point and the ith perturbed point.

A representative example of our numerical results was for example with the benchmark problem corresponding to the shape defines in the Cartesian coordinates by $(X(t), Y(t)) = (R_0\cos(t), -R_0\sin(t) + X(t))$. The inverse problem requires 1,200 forward simulation with 20 control points which is a fair amount of calculation. The relative error expressed in the L_2 norm on the coordinate of the control points was equal to 0.73%. The larger the number of control points the smaller should be this error. To solve efficiently this inverse problem in three space dimension will undoubtedly benefit from a parallel version of the sweeping algorithm that exploit the relative independence of distanced control points. This work is under development for 3D clinical cases that are necessary to provide a validation of the method. This reconstruction of the unloaded shape must be done only once. From that result a forward simulation with arbitrary direction of gravity can be done very quickly. However, as mentioned earlier, there is a fair amount of uncertainty on mechanical soft tissue properties with clinical situation. Next, We are going to explore the sensitivity of the forward simulation to material properties.

Influence of Tissue Description

We are mainly concerned by how important is the geometric distribution of soft tissue mechanical properties and what is the impact of internal stiffer tissue on the whole deformation.

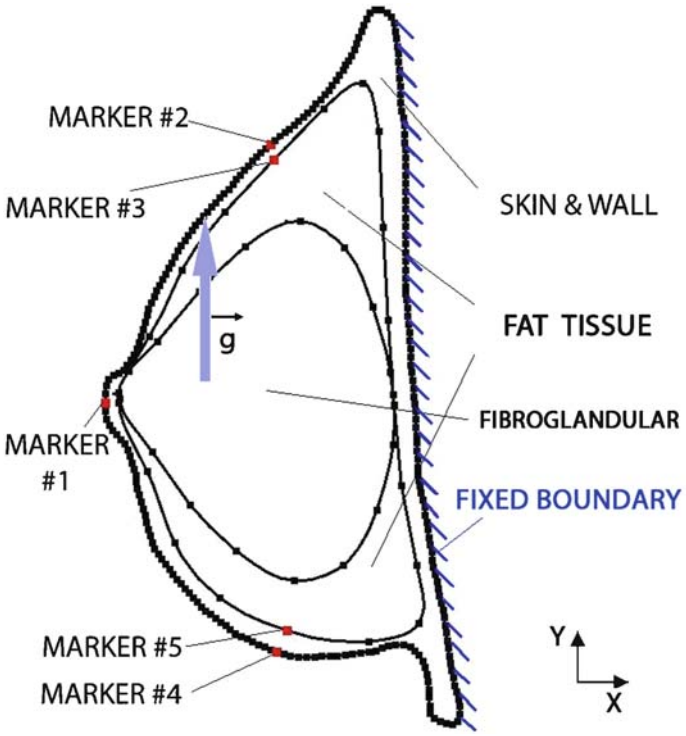


Fig. 8 Model for the sensitivity analysis of the fibroglandular tissue

Table 1 Young's modulus E of fatty and fibroglandular tissue for linear

Young's modulus E in Pa			
Name	Fatty tissue (E_f)	Fibroglandular tissue (E_g)	$\frac{E_f}{E_g}$
MM1	2,500	2,500	1
MM2	2,500	3,750	1.5
MM3	2,500	12,500	5
MM4	2,500	25,000	10
MM5	2,500	37,500	15
MM6	2,500	50,000	20

To tackle this question, let us consider a 2D breast model with a realistic distribution of tissue. The geometry of the breast, as well as the tissue distribution is given in Fig. 8.

We differentiate three types of tissue: fat, fibroglandular, skin. The individual tissue types were modeled as isotropic and homogeneous. Linear elastic model was used, and we tested a variety of Youngs modulus ratios for each tissue types as in Table 1), compatible with data find in the literature [24]. Six linear models were

constructed where fibroglandular tissue was 1, 1.5, 5, 10, 15 or 20 times stiffer than fatty tissue. These six models were tested without skin (MM1 − MM6), and with skin. The skin envelope was supposed to be four times stiffer than the fatty tissue in the linear models (MM1S − MM6S). We start from a loaded configuration with the person standing up. To test our model, we apply a body load $\| \overrightarrow{g} \| = 9.8\,\mathrm{m\,s^{-2}}$ in the upward $\overrightarrow{y}$ direction. This would correspond to the position of the breast at zero gravity. The deformation of the breast was monitored by plotting the displacement of five markers attached to the breast. Figure 9 illustrates the displacement of the five markers, placed on the breast model (see Fig. 8), along the $\overrightarrow{x}$ and $\overrightarrow{y}$ directions for the different Young's modulus values. As we can see, the general behavior is the same between the model with skin and the model without skin. The sensitivity of the model to the skin effect decays as the fibroglandular tissue is stiffer.

We notice as expected that the stiffer is the fibroglandular tissue, the smaller is displacement. There is however a saturation effect and we get roughly the same displacement when fibroglandular are 10, 15 and 20 times stiffer than fat.

Secondly, we have tested the effect of the presence of a small tumor on these deformations. We place some disk of 1-cm radius to model a tumor and look at the

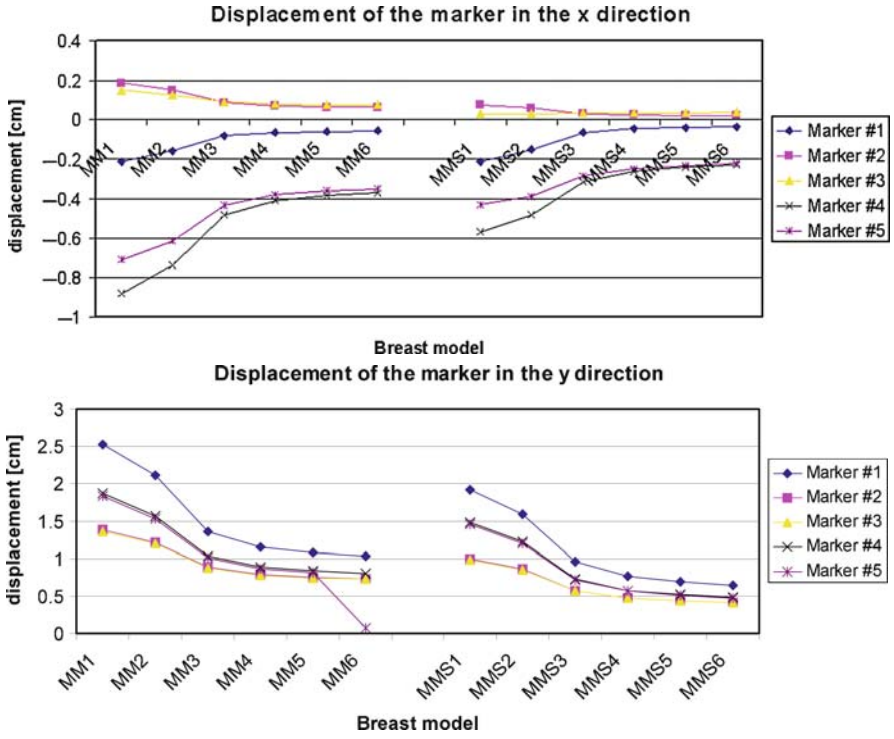


Fig. 9 Displacement of the different marker along x and y direction after loading

impact on the deformation of the breast while the tumor is supposed to be 25 stiffer than the fat tissue. We observe mainly a local perturbation of the previous results. In particular, the markers closest to the tumor get more affected as one can expect.

This sensitivity analysis reveals (with a two dimensional simulation), the importance of the choice of the breast model. An homogeneous tissue description to model the breast would be inaccurate. However, regarding the general behavior, an exact description of the tissue property is not mandatory. Moreover, this will lead to excessive constrain at the input of our model. One speculates that a coarse distribution of the tissue is needed to have prediction that fits within a few millimeter accuracy. Regarding the skin introduction in the model, the skin affects the amplitude of displacement, however the skin introduction does not impact the general behavior of the deformation.

With three dimensional clinical data, we expect to retrieve some good tissue composition indicator with either a coarse segmentations of MRI images, or even better elastography imaging. We can expect however some uncertainty left on those tissue mechanical properties, and a practical solution would be to calibrate the mechanical model with breast external 3D images obtained in various positions relative to the gravity. The solution of this identification problem may use the same optimization framework than the one we use for recovering the unloaded position.

Conclusion

The motivation of this study was to highlight the different aspects that it takes to develop a reliable mechanical model of breast deformation for BCT, in order to provide a good predictive tool for surgery. We discussed the choice of the model with linear elastic, hyperelastic approximation to simulate breast tissue and multimaterial description in particular for tissue removal. We found global deformation and local deformation due to BCT with good correlation with surgeons' observations. A sensitivity analysis was run to determine what were the most significant parameters to measure. We have shown that we can learn much from artificial benchmark problems and they should be used as a preliminary test of methods and ideas for BCT. However, a reliable predictive model needs much more work and should start from the clinical data we have started to accumulate. Our next step will be to concentrate on a mechanical model that retains enough of the desired characteristic to predict outcomes of BCT, but can still work in a clinical environment characterized by a significant level of uncertainty on tissue properties. Later on, this mechanical model will be coupled to a tumor growth model that should provide information on negative margin in surgery planning.

References

1. Benda RK et al (2004) Breast-conserving therapy (BCT) for early stage breast cancer. J Surg Oncol 85:14–27
2. Sivaramakrishna R (2005) 3D breast image registration – a review. Technol Cancer Res Treat 4:39–48
3. Azar FS, Metaxas DN, Schnall MD (1999) A finite element model of the breast for predicting mechanical deformations during interventional procedures. Proc Int Soc Magn Reson Med 7:1084
4. Sciaretta J, Bishop J, Samani A, Plewes DB (1999) MR validation of soft tissue deformation as modeled by non-linear finite element analysis. Proc Int Soc Magn Reson Med 7:246
5. Williams C, Clymer B, Schmalbrock P (1999) Biomechanics of breast tissue: preliminary study of force-deformation relationship. Proc Int Soc Magn Reson Med 7:524
6. Azar FS, Metaxas DN, Schnall MD (2002) Methods for modeling and predicting mechanical deformations of the breast under external perturbations. Med Image Anal 6:1–27
7. Rajagopal V, Chung JH, Bullivant D, Nielsen PMF, Nash MP (2007) Determining the finite elasticity reference state from a loaded configuration. Int J Numer Methods Eng 72:1434–1451
8. Krosukop TA et al (1998) Elastic moduli of breast and prostate tissues under compression. Ultrason Imaging 20:260–274
9. Manduca A et al (2001) Magnetic resonance elastography: non-invasive mapping of tissue elasticity. Med Image Anal 5:237–254
10. Rajagopal V, Lee A et al (2008) Creating individual-specific biomechanical models of the breast for medical image analysis. Acad Radiol 15:1425–1436
11. Wolfram S (1994) Cellular automata and complexity, collected papers. Westview Press, Boulder, CO
12. Holland JH (1995) Hidden order: how adaptation builds complexity. Helix Books, Reading, MA
13. Deutsch A, Dormann S (2005) Cellular automaton modeling of biological pattern formation. Birkhäuser, Boston
14. Hwang M, Garbey M, Berceli SA, Tran Son tay R, Ruled-based simulation of multi-cellular biological systems - A review of modeling techniques. Cell Mol Bioeng, in press
15. Bailey AM, Thorne BC, Peirce SM (2007) Multi-cell agent-based simulation of the microvasculature to study the dynamics of circulating inflammatory cell trafficking. Ann Biomed Eng 35:916–936
16. Mallet DG, De Pillis LG (2006) A cellular automata model of tumor-immune system interactions. J Theor Biol 239:334–350
17. Mi Q, Rivire B, Clermont G, Steed DL, Vodovotz Y (2007) Agent-based model of inflammation and wound healing: insights into diabetic foot ulcer pathology and the role of transforming growth factor-β1, Wound Repair Regen 15:671–682
18. Galle J, Hoffmann M, Aust G (2009) From single cells to tissue architecture-a bottom-up approach to modelling the spatio-temporal organisation of complex multi-cellular systems. J Math Biol 58:261–283
19. Thorne BC, Bailey AM, Peirce SM (2007) Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. Brief Bioinform 8:245–257
20. Berceli SA, Tran-Son-Tay R, Garbey M, Jiang Z (2009) Hemo-dynamically driven vein graft remodeling: a systems biology approach. Vascular 17:S2–S9
21. COMSOL 3.4, COMSOL Multiphysics. COMSOL Reaction Engineering Lab
22. Schwartz J-M (2003) Calcul rapide de forces et de deformations mecaniques non-lineaires et visco-elastiques pour la simulation de chirurgie. PhD, University Laval
23. Nelder JA, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313
24. Tanner C, Schnabel J, Hill D et al (2006) Factors influencing the accuracy of biomechanical models. Med Phys 33:1758–1769
25. Ribba B, Colin T, Schnell S (2006) A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7

Part V
Training

Simulators in Training

Brian J. Dunkin, MD, FACS

Introduction

The rapid pace of change in medical technology has put unprecedented pressure on clinicians to learn new procedures and instrumentation and adopt them safely into their practice without compromising patient outcomes. No longer can a surgeon rely on the skills learned in training to carry them through their career. In the last two decades alone, medicine has seen the introduction of video cameras, ultrasound machines, CT scanners, MRI machines, and robots into the operating room in an effort to provide less invasive and more exact surgical care. Learning to use these devices safely and rapidly is a constant challenge. In addition, cost constraints and concerns about patient safety make it no longer acceptable to practice procedural technique in the real clinical environment. Finally, medical procedures have advanced beyond a "one man show" requiring the careful coordination of multiple team members to be successful.

All of these challenges have caused the medical community to turn to the use of simulators to create a noncritical environment in which to rehearse patient care and gain feedback about performance. The types of simulators used vary greatly in sophistication and design, but can be categorized into three groups – inanimate, animate, and cadaver. This chapter will focus on the development and use of inanimate simulators. The different types of inanimate simulators will be described along with their design. The development of metrics of skills acquisition will also be described and a clinical example of designing a surgical simulator, creating metrics of skills acquisition for it, and proving that practice on simulated tasks improves real clinical performance will be given.

B.J. Dunkin (✉)
Department of Surgery, Weill Cornell College of Medicine, The Methodist Hospital, Houston, TX, USA

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_16,

Types of Inanimate Simulators

Inanimate simulators can be categorized as partial task, full task, or human patient simulators with or without computer assistance. An example of a partial task simulator is the airway mannequin shown in Fig. 1. This model mimics the anatomy of a patient's head, oral cavity, upper airway, and neck and is used to teach clinicians how to insert a plastic tube into the trachea to assist a patient's breathing – a procedure called airway intubation. It is a partial task simulator, because it does not simulate all of the steps required to successfully intubate such as, the proper administration of sedation, bag ventilation prior to intubation, and listening for breath sounds after intubation. A full task simulator is shown in Fig. 2. This is a computer-driven simulator that recreates the hospital angiography suite and is used to learn how to manage complex vascular problems by delivering devices inside the lumen of the blood vessel without doing invasive surgery – a process called endovascular surgery. The simulator recreates all parts of the procedures, including creating a sterile environment, properly administering medications, accessing the correct blood vessel, delivering devices to an area of disease, and preventing and managing complications of the treatment. If the patient is not treated correctly, she may actually "die" on the table! Figure 3 shows a human patient simulator (HPS). These sophisticated mannequins are also computer-driven and designed to mimic a whole human with different disease states. The HPS blinks, breathes, talks, and secretes. Blood samples can be drawn from it, catheters inserted, and medications delivered with the appropriate physiologic response. Such simulators are frequently used in team

Fig. 1 A partial task simulator created from three-dimensional imaging of the human mouth, throat, and upper airway used to practice intubation (AirSim, Trucorp, Ltd. Belfast, N. Ireland)

Fig. 2 A full task simulator for practicing endovascular procedures. The simulator recreates the entire radiology suite (SimSuite, Medical Simulation Corporation, Denver, CO)

Fig. 3 A human patient simulator capable of mimicking human physiologic states (iStan, Medical Education Technologies, Inc. Sarasota, FL)

training for low frequency, high risk events such as a heart attack. It is not common for a patient to "code" in the hospital with an acute loss of pulse, blood pressure, and breathing, but when it does happen, you want a resuscitation team that is well rehearsed in managing this life threatening situation. An HPS can be programmed to mimic a "code" and then the resuscitation team can rehearse their response until satisfactory performance is achieved.

Procedural Simulation

The rapid pace of change in medical technology has put great pressure on clinicians to constantly learn new procedures to stay current. As a practicing physician, it is challenging to leave your practice for a period of time to learn a new procedural

technique or technology and then bring this procedure back to your patients in a safe and efficient manner. As an instructor for the new procedure, it is difficult to know if the trainee has mastered the skills required to perform the procedure safely without watching him or her work on real patients in their own hospital – a proctoring process that has become nearly impossible to do in today's medical-legal climate. As a result, physicians are increasingly looking at using simulators to recreate the clinical environment and allow rehearsal of a procedure in a safe and cost-effective manner. This concept is not new. Animals and cadavers have been used for centuries to study anatomy and rehearse procedures. Simple inanimate models have also been used to learn basic tasks, such as suturing and knot tying. Almost every medical student interested in surgery has spent time tying suture knots onto his or her clipboard or lab coat – a simple but effective simulator for sure! The advent of powerful personal computers has enabled simulator designers to generate high fidelity graphics, drive force-feedback mechanisms, and follow treatment algorithms to increase the realism of simulation. However, despite the development of technologically sophisticated simulation platforms which appear to mimic real life anatomy or physiology, it has been difficult to show that these procedurally based simulators actually improve clinical performance. In colonoscopy, for instance, despite training for multiple hours on a sophisticated computer-driven simulation platform costing tens of thousands of dollars, two studies showed that the experience only helped the trainee perform better than the nonsimulator trained trainee during the first 20–80 cases and did not decrease overall the number of procedures required to meet proficiency [1,2]. The reason for these kinds of results is because most procedural simulators are designed backwards. The simulator is created as a computer game to look as much as possible like the real clinical environment. Talented computer programmers use images reconstructed from three-dimensional X-rays to recreate the look and feel of surface anatomy. Expert clinicians are used to confirm that the simulator indeed looks like the real clinical environment. However, only after this process is completed does work get done to test weather or not training on the simulator actually improves clinical performance.

The better way to design a simulator is to start with determining what is important to performing the desired clinical task. Experts must deconstruct this task into specific skill sets that one must possess to do the procedure well. The simulator should then be designed to recreate these specific skills sets and a system of measurement devised to determine if the trainee is performing the skill well. Finally, a valid assessment tool of live clinical performance must be devised in order to determine if training on the simulator improves a clinician's ability to perform the procedure in the real world.

Currently, in surgery there is only one commercially available simulator that has been designed and validated using the process described above. In order to more fully illustrate the required steps to design such a simulator and deploy it within a training and testing curriculum, the process of developing and testing this simulator will be reviewed.

Developing a Laparoscopic Surgery Simulator

In 1989, Dr. Eddie Joe Reddick became the first surgeon in the US to remove a diseased gallbladder from a patient using a miniaturized camera system, long narrow instruments, and four half inch incisions in the abdominal wall [3]. This new technique was called laparoscopic surgery and it would soon revolutionize abdominal and thoracic surgery forever. Within 3 years, nearly all general surgeons in the US would move from performing gallbladder surgery using an open technique – create a large scar under the rib cage and use direct visualization of the internal organs – to laparoscopic surgery. Most had never even seen a laparoscopic procedure before, much less used the technique to perform a complex operation. This created a demand for new procedure training in surgery that had never been seen before. Nearly overnight practicing surgeons needed to learn an entirely new set of surgical skills and offer them to their patients in a safe manner. As predicted, there was significant variation in the types of training experiences offered and, unfortunately, some patients suffered complications from undergoing operations by inadequately prepared surgeons.

In an effort to create a standardized training platform in laparoscopic surgery, a group of surgeons at McGill University in Montreal, Quebec, Canada decided to build an inexpensive, portable simulator to measure skill in performing laparoscopic surgery [4]. They felt that it was important to not only design a simulator that mimicked the desired skills, but to also develop a set of measurements (metrics) that could be used to show that a trainee had achieved proficiency on the skills. The result of their work was the MISTELS (McGill Inanimate System for Training and Evaluation of Laparoscopic Skills) physical laparoscopy simulator.

The MISTELS simulator consists of five exercises performed in a trainer box. A camera and LED lighting system provide visualization and the surgeon works through two trocars placed in fixed positions through an opaque cover on top of the box (Fig. 4). The system was designed to be inexpensive and easily portable so that it could be used at home for practice.

To develop the five standardized tasks used in the trainer box, five experienced laparoscopic surgeons reviewed video tapes of various laparoscopic procedures and listed the differences between performing the surgeries using laparoscopy versus open surgery. The principle differences identified were use of an optical system which provided magnified monocular vision that affected depth perception, use of long instruments that diminish tactile feedback and amplify tremor, and the need to work through trocars (air tight plastic tubes inserted through the abdominal wall) with fixed access, decreased degrees of freedom, and a resultant fulcrum effect with the instrument pivoting at the abdominal wall causing the tip to move in the opposite direction of the hand controlling the instrument [4]. In addition, specific devices unique to laparoscopic surgery were identified to ligate structures and tie suture and these devices must be directed to their target with fine two-handed motion. Based on this "deconstructed" list of skills required to perform laparoscopic surgery, five specific tasks were devised for training and evaluation and are illustrated in Fig. 5.

Fig. 4 A MISTELS-type box trainer (Inanimate Skills Laboratory, Methodist Institute for Technology, Innovation, and Education, Houston, TX)

After devising the simulator platform and tasks, it became important to develop metrics where each exercise is scored for efficiency and precision. For the MISTELS simulator, a cutoff time was assigned for each task. A time score as a measure of efficiency was then calculated by subtracting the actual task time from the cutoff time. A penalty score was applied for errors or lack of precision and the penalty score is subtracted from the efficiency score for each task. The five task scores were then normalized to create an equivalent range and added for a total score.

Determining the Reliability and Validity of a Simulator

If a simulator is to be an effective educational tool, the metrics associated with the tasks it simulates must provide meaningful information to the trainee and be shown to be reliable and valid. The MISTELS simulator has gone through vigorous reliability and validity testing and will be used as an example to illustrate this process.

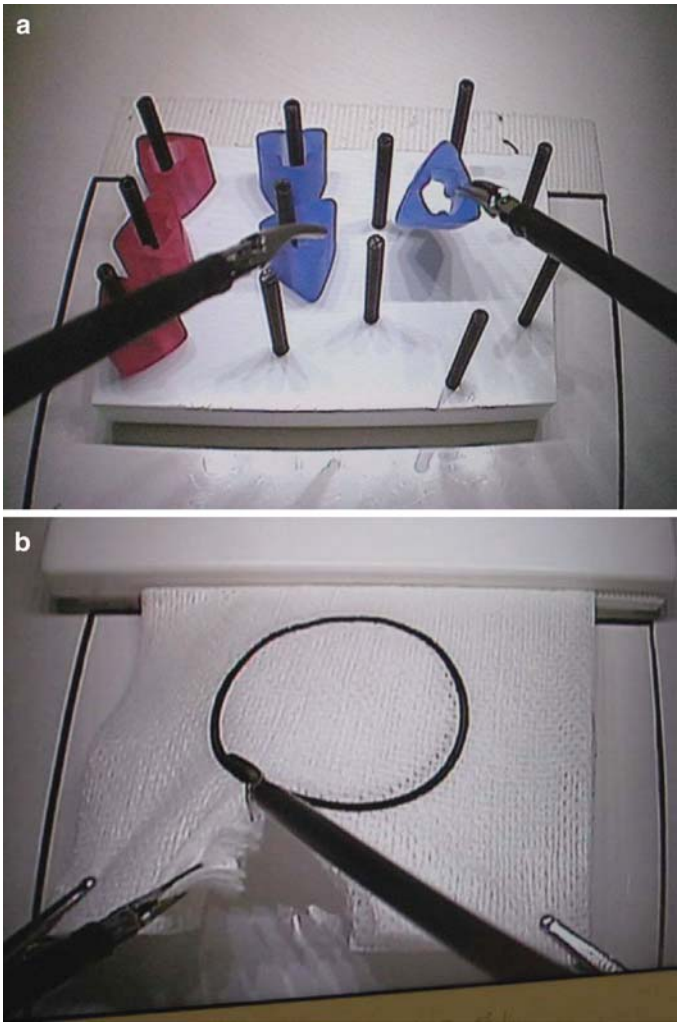


Fig. 5 MISTELS tasks (**a**) Peg transfer: pick up each colored object with *non-dominant* hand and transfer in mid-air to *dominant* hand and place on the other side. (**b**) Pattern cut: cut out a *circle* directly on the black line, balancing accuracy with time. (**c**) Endoloop: place the ligating loop precisely on the black line, secure and cut. (**d**) Suturing and knot tying: introduce the needle, load on the instrument, and aim for the dots to close the gap in the drain by performing three throws of a square knot

Reliability refers to the consistency of the test. It must result in similar scores when administered by different testers (inter-rater reliability) and must be consistent when the same trainee is evaluated on different occasions with no new learning between tests (test–retest reliability). It must also be internally consistent, meaning the individual task scores mirror the overall score. The MISTELS simulator has

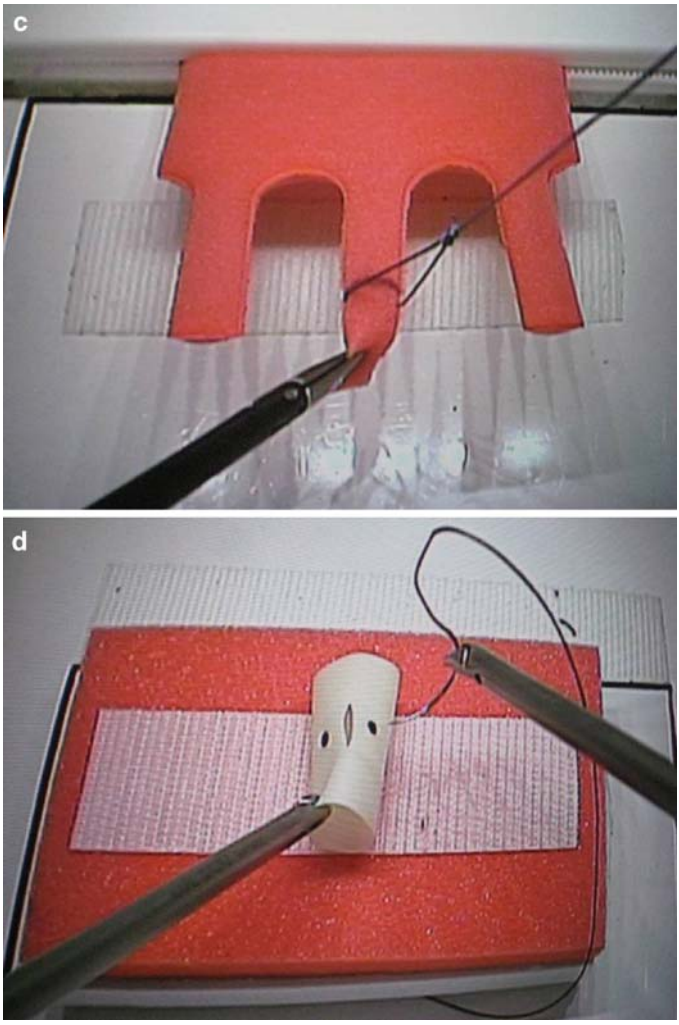


Fig. 5 (continued)

been proven to have excellent interrater and test–retest reliability as well as internal consistency [4]. Demonstration of such reliability is mandatory before an evaluation can be shown to be valid.

Once reliability of measuring performance on a simulator is shown, proof of validity is then established by conducting a series of independent studies to ensure that measurements in the simulator are actually measuring the desired skill. Five types of validity must be established – face, content, construct, external, and criterion.

Face validity asks does the simulation resemble the real task? It is usually established by having a group of experts in the procedure work on the simulator and then

complete a questionnaire which asks if the simulator includes all the skills required to complete the desired task. Face validity should not be confused with making the simulator look like real life. The MISTELS simulator uses a game-like environment that looks nothing like performing a real laparoscopic operation, yet experts scored it highly on face validity because the five tasks obviously encompassed the skills required to perform laparoscopic surgery.

Content validity asks if the simulation measures all relevant dimensions of the desired procedure. It is also established by having a group of experts perform the simulated tasks and report via a questionnaire if they believe any required skills are missing. Overall, the MISTELS simulator performed well in content validity when exposed to a group of expert laparoscopic surgeons although there were one or two additional tasks that could have been included and work in this area continues for the next generation of simulator.

Construct validity asks does the simulator performance score agree with another construct that can be reasonably assumed to correlate with technical skill – in this case in the operating room. A common hypothesis used in this area of validation is that experts should score better than novices on the simulator. MISTELS was used to test over 200 surgeons of varying backgrounds and levels of experience and found to have construct validity with experts scoring better that intermediately experienced surgeons who scored better than novices (Fig. 6).

External validity assesses the generalizability of the scores achieved on the simulator. Ideally, the scores should not vary significantly from test site to test site provided the exam is administered by a qualified proctor. External validity was confirmed for the MISTELS scoring system by having trained evaluators at five different test sites administer the exam. Scores were then compared to experienced

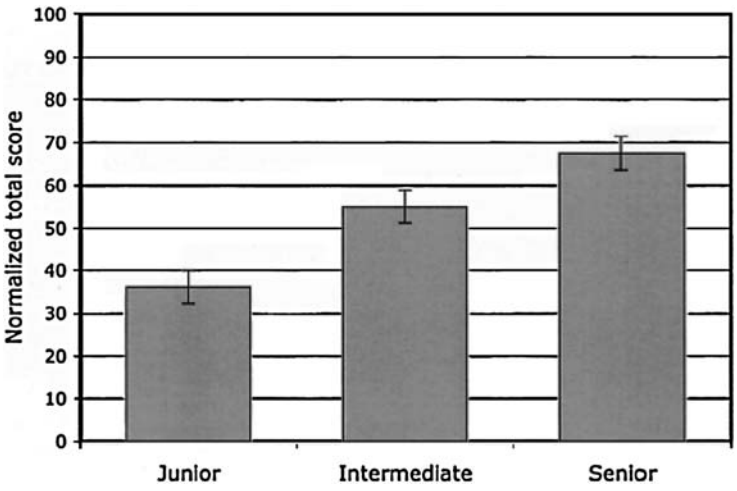


Fig. 6 Comparison of MISTELS total scores among junior residents (1st and 2nd year), intermediate residents (3rd and 4th year) and seniors (5th year, fellows, and practicing surgeons). $N = 215$, $p < 0.0001$ (Reprinted with permission from ref. [4])

matched groups at McGill University where construct validity was established for MISTELS. This comparison showed that only the level of training and not the test site independently predicted the total MISTELS score.

Criterion validity is divided into two areas – concurrent and predictive. Concurrent validity asks if performance on a task on the simulator mirrors performance in real life. Predictive validity asks if performance on the simulator will predict future performance. Predictive validity is the "Holy Grail" of simulation training because it establishes that performance on the simulator predictably measures performance in the real clinical environment. A number of simulation platforms have been shown to have face, content, and construct validity, but very few have predictive validity.

Concurrent validity for MISTELS was established by comparing performance on the simulator to another measure of clinical skill. The first comparison was to performance of similar tasks developed and measured in a live animal model [5]. Second was a comparison to the in-training technical skill evaluation scores of surgery residents [6]. In both cases, performance on the MISTELS simulator correlated with performance on the independent measure of clinical skill.

The most important validation of the MISTELS trainer was to show that it had predictive validity – i.e., that performance on the simulator will predict performance in the operating room. To prove this level of validation required a critical measurement tool that had not been developed up to this point – a valid test of laparoscopic skills in the real operating room. To address this need, the Montreal group developed the Global Operative Assessment of Laparoscopic Skills (GOALS) [7]. This is a global assessment score that can be used to measure how well a surgeon performs laparoscopic surgery in the operating room. It has been proven to be reliable and valid which separates it from almost all clinical assessment tools of procedural competence. The MISTELS simulator scores correlated highly with the intraoperative measurement of technical skills using GOALS (Fig. 7). Thus, the MISTELS

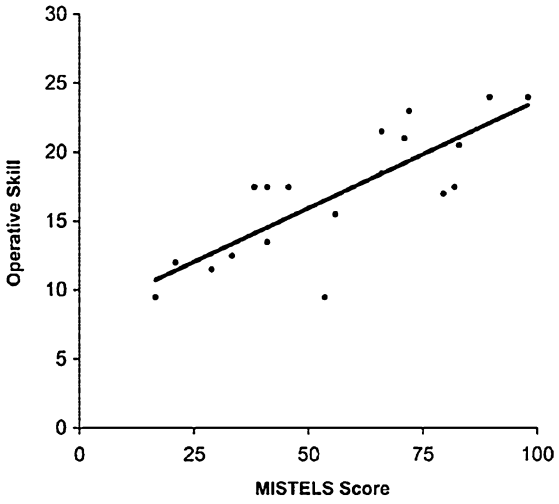


Fig. 7 Linear regression analysis of operative skill measured by a validated global assessment scale during laparoscopic cholecystectomy versus MISTELS score. $N = 19$, $r = 0.81$, $p < 0.001$ (Reprinted with permission from ref. [4])

simulator was proven to have predictive validity meaning that those who performed well on the simulator could be expected to perform well in the operating room.

This important step in the validation of a procedural surgical simulator had never been accomplished until this work. For the first time, it was reliably proven that performance on a surgical simulator could predict performance in the real operating room.

Using Simulation to Test Procedural Skill

Once a simulator has been shown to be reliable and associated with valid metrics of performance, it can then be used as an assessment tool. To do this, the sensitivity and specificity of performance on the simulator needs to be described. For the MISTELS simulator, receiver operator curves were generated and showed that a cumulative score of 270 maximizes sensitivity, specificity, and positive and negative predictive values (each $>$ 0.80) [8]. Thus, the pass/fail cutoff score for MISTELS was set at 270 (Fig. 8).

Surgeons recognize, however, that becoming an expert requires knowledge and judgment in addition to technical skill. As a result, a comprehensive assessment tool for laparoscopic surgery must test both knowledge and skill. With the MISTELS simulator now available to reliably test technical skill in laparoscopic surgery, the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) created a cognitive exam which has also gone through rigorous reliability and validation testing. The two tests were packaged together into a program called the Fundamentals of Laparoscopic Surgery (FLS) and made available to all physicians who wish to be tested.

Cutoff score	NC % PASS	C % PASS
240	28	89
250	27	89
260	23	84
270	18	82
280	16	79
290	13	76
300	12	74
310	11	73
320	6	71
330	5	66
340	2	61
350	1	59

NC, noncompetent; C, competent

Fig. 8 The MISTELS simulator, receiver operator curves show that a cumulative score of 270 maximizes sensitivity, specificity, and positive and negative predictive values (each >0.80).[8] Thus the pass/fail cutoff score for MISTELS was set at 270

FLS is the first commercially available validated test of surgical knowledge and skills. It has started a firestorm of demand in multiple areas. In Massachusetts, surgeons in one hospital system who perform laparoscopic surgery receive a lower rate on their malpractice insurance if they pass FLS. In the US military, surgeons performing laparoscopic surgery are also receiving FLS testing and certification. And in 2009, FLS received its most important endorsement when the American Board of Surgery (ABS) began requiring FLS certification for all general surgeons wishing to become ABS certified.

Conclusion

The rapid pace of change in modern medicine makes training on simulation platforms an absolute necessity. Unfortunately, the current state of development for simulated procedural environments remains rather limited and the availability of validated metrics of skills acquisition is just beginning. When developing a simulation platform, it is paramount that work proceeds in the correct order. It should not start with developing a gaming environment which appears to mimic the real clinical scenario. Instead, work should begin by deconstructing the desired task into necessary skills sets and then developing the simulator to test those skills. A clear set of metrics must then be developed to ensure that skills acquisition can be measured in an objective fashion, and a reliable and valid test of real clinical performance must also be devised. Then, and only then, the simulator must go through rigorous testing for reliability and validity. Once this is finally established, the simulation platform can be used to practice or as part of a testing platform to demonstrate facility in a particular technical skill.

If this pathway to simulation development is followed, medical training and credentialing will evolve to a point where a surgeon will not be allowed to independently perform a new operation or procedure without first rehearsing in a simulated environment. These environments will continue to become more sophisticated as well until patient specific simulation becomes commonplace. Thus, in the future, when a patient enters the operating room for a surgical procedure, the surgeon will not only have trained to do that procedure on a validated simulation platform, but also rehearsed the procedure using the patient's own physiologic and radiologic data to make the experience as specific to that individual's disease process as possible.

References

1. Sedlack RE, Kolars JC (2004) Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 99(1):33–37
2. Cohen J, Cohen SA, Vora KC et al (2006) Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointest Endosc 64(3):361–368

3. Reddick EJ, Olsen DO, Daniell JF et al (1989) Laparoscopic laser cholecystectomy. Laser Med Surg News Adv 7(1):38–40
4. Fried GM, Fledman LS, Vassiliou MC et al (2004) Proving the value of simulation in laparoscopic surgery. Ann Surg 240(3):518–528
5. Fried GM, Derossis AM, Bothwell J et al (1999) Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator. Surg Endosc 13:1077–1081
6. Feldman LS, Hagarty SE, Ghitulescu G et al (2004) Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg 198:105–110
7. Vassiliou MC, Feldman LS, Andrew CG et al (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 190(1):107–113
8. Fraser SA, Klassen DR, Feldman LS et al (2003) Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system. Surg Endosc 17:964–967

A Computational Desk for Surgeons

Victoria Hilford, Yusuf Yildiz, and Marc Garbey

Problem and Motivation

According to the American Heart Association, cardiovascular disease is an underlying cause of death that accounted for 37.3% of all deaths, or one of every 2.7 deaths in the United States in 2003. Many of these deaths corresponding to vascular disease are due to the rupture of an aneurism or the rupture of a vulnerable plaque usually located near the narrowing of blood vessels, also called stenosis [1, 2]. The interested reader can refer to the chapters of Mark Davies et al. and Scott Berceli et al. in this volume for an overview of low extremity vessels diseases.

Endovascular surgery is a relatively new procedure to treat such problems that may develop in blood vessels. In recent years, we have seen the development of intravascular balloons, stents, and coils that can be put in place through arteries with minimum invasive procedures. Endovascular surgery is performed by radiologists, neurosurgeons, cardiologists, and vascular surgeons. This surgery is essentially guided by imaging such as X-ray techniques. The interested reader can refer to the chapter of Mark Davies et al. for a review of the state of the art, and the chapter of Berceli et al. that emphasize the system biology approach of the problem.

Typically, workflow involved in the planning of endovascular surgery includes the acquisition of the patient's medical image in DICOM (Digital Imaging and Communications in Medicine) format, the image visualization such as zooming, rotating, 2D and 3D rendering that allows the surgeon to identify a region of interest (ROI), the image processing that allows segmentation of the original medical image based on the chosen ROI, creating a model by using the geometry extraction on the ROI, extracting patient-specific boundary condition parameters for flow simulation, running the simulation, and visualization of the results. The surgeon will repeat these steps easily as many times as needed to come up with a decision. The surgeon will also be able to virtually modify the geometry of artery lumen, in order to anticipate the outcome of a surgical procedure.

V. Hilford (✉), Y. Yildiz, and M. Garbey
University of Houston, Houston, TX, USA
e-mail: vhilford@cs.uh.edu,yildizyusuf@gmail.com,garbey@cs.uh.edu

M. Garbey et al. (eds.), *Computational Surgery and Dual Training*,
DOI 10.1007/978-1-4419-1123-0_17,

Currently, several tools exist that partially integrate some of the steps in the typical workflow described earlier. We are interested in building an integrated tool that can integrate clinical information with modeling and simulation as reviewed in the chapter by Mark Davies et al.

The difficulties are multiple and occur at all levels from data representation, to data processing, to human computer interfaces. The data manipulated for imaging, simulation or simply medical administration are completely heterogeneous and may have evolving formats. There is a high level of uncertainity due to the nature of medical exams and medical image interpretation. In practice, one faces a lack of robustness of hemodynamic simulation, i.e., it takes more than one simulation to verify the result. Further, the elapsed time required to process fluid dynamics is considerably larger than for image processing. We tried to address these various challenges in our design.

The difficulty in providing a friendly computer interface to surgeons for accessing all the data of interest at once on multiple displays should not be underestimated. The interface should present the data in a hierarchic way with iPhone-like buttons that will give the surgeon access to details only if desired. The graphic user interface (GUI) should be as simple as possible, but should not limit the spectrum of information available.

The quantity of data to be accessed is huge. We are dealing with a large spectrum of image modalities and numerical formats. The system should provide fast access to the data because surgeons have limited time to decide on the procedure. The interface must be friendly and very intuitive for the surgeon. Finally, there are additional requirements such as ensuring that the patient data are kept private (HIPAA – Health Insurance Portability and Accountability Act) and the system is reliable.

Our computational desk for surgeons (CDS) proposes an integrated solution of the typical endovascular workflow. In addition, it provides a database that stores not only the original medical images, but also the intermediate data together with the annotations that the doctor has entered through the interface. The database user interface allows the surgeon to recall old patients data for possibly redoing the workflow now based on a medical dataset from the database. It is also important to keep track of the history of the decision process for liability purposes.

The integrated individual subsystems, namely image visualization, image processing, simulation, visualization, and database are all freely available to some extent. The novelty of the CDS system is combining all the endovascular workflow steps into a single integrated solution. The incompressible Navier–Stokes solver in particular is an innovative fast parallel and robust algorithm that interfaces naturally with level set image processing thanks to an immersed boundary technique. This paper summarizes and adds on previous work we did on the CDS project [3, 4].

In section “Review of Existing Systems,” we will briefly review the literature on some of the critical components in medical imaging and flow processing. Section “The System Architecture and Workflow Scenarios for Endovascular Surgery Case Study” presents our system architecture and workflow scenario for a benchmark endovascular surgery case study. In section “Conclusion,” we will summarize our result and comment on our future work.

Review of Existing Systems

We are going to briefly review three broad areas that contribute in providing a practical solution for endovascular surgery. We will concentrate on few main contributions published in these areas and we will not be able to do proper justice to all other good contributions published.

First is the patient management area. A surgeon has to comply with his institutional work environment and the healthcare regulations to manage patients. Every day, physicians have to write notes about their clinical examinations. They want to have a trace of their diagnosis and observation, for further information retrieval or simply to keep track of the case. Indeed, some physicians use handwritten notes, but they could also use another person like a scribe to write in their place, and then dictate letters. But for several years, the annual cost of medical transcriptions is about 10 billion and can reach up to 24 billion dollars. In addition to a high cost, transcription takes up a lot of time of physicians and other people who work on it [5, 6]. With an increase in electronic health record (EHR) use, most of patient data are digital. Another way for physicians to enter and edit clinical notes about patients is speech recognition [7]. This technology could make the medical transcriptions more productive and the potential for saving is enormous [8]. We found at least two references of systems that use speech recognition software for medical applications. The first, using Dragon Naturally Speaking Medical Suite, is Medical Speech Interface System (MSI) developed by the researcher from UMDNJ-SHRP Newark, NJ. MSI system allows the end user to input sets of clinical data via voice [9]. The second system, from University of Amsterdam in the Netherlands, includes speech recognized user command interface into their application [10]. One of the typical patient information that is manipulated is indeed medical images. Image annotation is becoming increasingly important in the medical world. Especially, collecting intelligence from distributed users can lead to better prediction and analysis based on medical images. One system that addresses the collaborative annotation was developed at Siemens Corporate Research [11]. They developed a software tool allowing to separate different users from each other, keeping the history of the annotations, enabling graphical annotations, and effective querying on the annotations accomplished by using XQuery. Another tool for image annotations is Ipad that follows the AIM standard [12, 13]. Finally, all these patient medical records have to be manipulated transparently at the level of the hospital infrastructure which need to communicate the information from one department like radiology to another like surgery or pathology in order to facilitate the collaboration between the various actors of the medical process. In this complicated and heterogeneous environment, patient data have to comply to the healthcare regulation and ethics, which make healthcare information technology very special. We refer in particular to the health insurance portability and accountability act – 1997 (HIPAA).

Second is the medical imaging and processing area. Medical images are of paramount importance in surgery. Images need to be acquired, stored, visualized, and interpreted in the most efficient way. There are a lot of commercial and non-commercial products for image visualization. Dicomworks is a standalone and

free dicom utility and viewer for desktop users [14]. Digital Jacket Pro is also a standalone commercial product which has strong features such as including a server to archive all the medical images and providing good search capability in the database [15]. Tera Recon is a technology solutions provider of advanced 3D imaging systems for medical world. Their image processing technologies focus on high performance reconstruction engines, real-time volume rendering, network-attached and real-time diagnostic workstations [16]. Some others can be listed as CoActiv, Biotronics 3D, Emageon [17–19]. Commercial applications are quite expensive and free utilities are hard to integrate. However, ImageJ provided by the National Institutes of Health, is an open-source, freely available, full-featured, and user-friendly program for image analysis. Moreover, ImageJ provides all the source codes. By plug-in development, someone can easily develop his own plug-in and start using it [20]. Consequently, ImageJ is used by many researchers to develop medical applications [21–23].

Modeling involves the determination of selected geometric features. Volumetric datasets capture big areas including unnecessary information which is not always diagnostically beneficial. Thus, modeling starts by choosing a ROI from the original medical image. A proposed system does modeling by ROI coding which enables faster reconstruction of diagnostically useful regions in volumetric datasets [24]. Researchers from Georgia Institute of Technology employ voxel classification and active surface techniques to extract and visualize the portal veins [25]. In biomedical application, one commonly used visualization with simulation. This trend started from the academic work of research team such as for example the center for integrative biomedical computing at the University of Utah headed by C. Johnson. Several software produced by this institute illustrates that effort with BioPSE for biomedical computing and visualization and VisTrails, that provides data and process management support for exploratory computational tasks [26].

While we have seen an amazing growth of tools in image processing and visualization, the interpretation of medical image is still an art where the experience of the medical doctor makes a huge difference.

Third is the application of the computational fluid dynamics simulations in medical care. Simulation in hemodynamic is used to show quantities that cannot be seen by medical imaging, at least with enough precision. Quantities of interest are shear stress at the artery wall, pressures, zone of flow recirculation, mechanical forces on stent, etc. One may eventually simulate exchange of substance between the lumen and the artery wall critical in atherosclerosis or drug delivery and possibly long term effect involved in tissue plasticity.

Imaging-based computational fluid dynamics (CFD) blood simulations have been used to analyze the flow pattern inside the important vessels such as aorta and coronary arteries, and this simulated information can aid the diagnosis and pre-surgical preparation. There are many research projects that use CFD, see for example [27–29]. Medis is a company located in Netherlands and provides wide range of measurement applications specializing in three different areas heart, vessels and brain [30]. Sim-Vascular and BloodSim are similar applications [22]. Mason Researchers developed a system focused on diagnostics and treatments of

brain aneurysms [31]. This volume also contains four chapters – see chapter by Berceli, Davies, Karmonik, Bismuth et al., that give a representative example of hemodynamic simulation applied to endovascular diseases based on a well-known commercial package named Fluent from the group Ansys [32].

Beyond commercial applications that are mostly based on the finite element theory, there are many variants in the world of numerical techniques to approximate the Navier–Stokes flow. Higher order methods with spectral elements have been developed by Paul Fischer and coworkers for hemodynamic studies. One of the great advantage of these high order methods is to catch phenomenon that are very sensitive to transition to turbulence and can be missed by lower order finite element techniques [33–35]. On the down side, these methods are extremely time consuming.

Powerful finite element numerical codes have been extensively developed to represent complex realistic hemodynamic flow with various degrees of complexity including fluid structure interaction, sophisticated techniques to connect local three dimensional simulation to all body network hemodynamic, etc. We refer to the work of Taylor and coworkers that provides an excellent example of this top of the edge technology [36–39].

A general observation on the state of the art of hemodynamic simulation is that it takes hours to run a simulation, and possibly days to go from raw medical images to final results. Meshing complex geometry of lumen is not that trivial, each simulation case may require special tuning, one single simulation is not that meaningful, because flow conditions change with body position or metabolic conditions. In other words, hemodynamic is by far not at the level of image processing in terms of elapse time. This overhead due to simulation may not be acceptable to incorporate the flow data into the pretreatment planning process. This is why most hemodynamic work is actually focused on research projects that try to establish bridges between mechanical forces and diseases. We will present in our design an innovative integrated approach to compute quickly an incompressible Navier–Stokes (NS) flow in a section of a large blood vessel using medical imaging data, that may bypass this difficulty.

Now, we are going to discuss the design of our CDS system.

The System Architecture and Workflow Scenarios for Endovascular Surgery Case Study

In the previous section, we presented a brief review of the workflow involved in endovascular surgery planning: medical image visualization, processing, modeling, simulation, and database storage. While some systems integrate a few aspects of this medical workflow, none attempt to integrate all aspects of the workflow. The computational desk for surgeons (CDS) is a unique system, to our knowledge.

We went conceptually through three generations in the design of our system. First, we started with a distributed approach, that combines powerful complementary and specialized resources acquired through an NSF equipment grant, see Fig. 1.

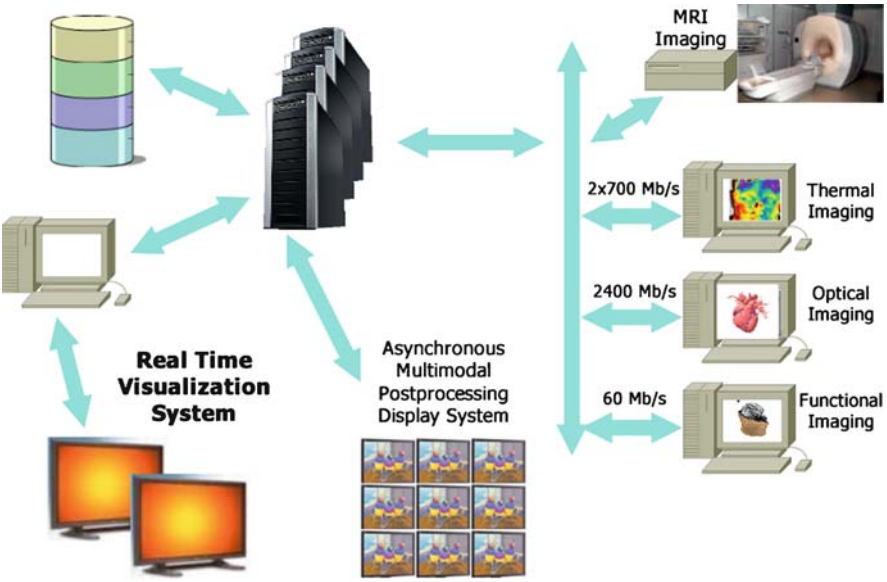


Fig. 1 First generation: MRI desk

It's advantage was having a system that many users could access. To maintain this system however required a high level system administrator, some specialized space devoted to the equipment and sharing between users who had different constraints. It was not clear that this distributed system was a cost effective solution either that can scale up. Our second generation system was a standalone solution with four fat water cooled PCs equipped with high end graphic cards and hanging below an Ikea desk with a reinforced frame. We found this solution cost effective, very efficient for processing and easy to maintain. The system was customized to one single user, and there were no issues on CPU time sharing. Figure 2 shows the IDV Desk in action. In this chapter, we are going to present our third generation system design that is more compact and user friendly.

Figure 3 shows the CDS system that we are currently assembling. Our goal is to improve on both cost effectiveness and ergonomic design.

We have modified a standard computer desk and used four water-cooled workstation towers. These workstations are mounted on supports hanging below on each side of the desk. An array of two rows by two columns of 22″ wide screen monitors are placed on the table in such a way that their orientation can be adjusted by the user. The lower monitors are touchscreen displays. This setup allows an effective usage of the CDS system.

Each workstation is based on multicore processors on each workstation with 8 GB of RAM. With the idea of performance computing and parallel processing in mind, we interconnected the workstations by a Gigabyte Ethernet switch. Three of the PCs are used for parallel simulation as well as parallel visualization. Each of

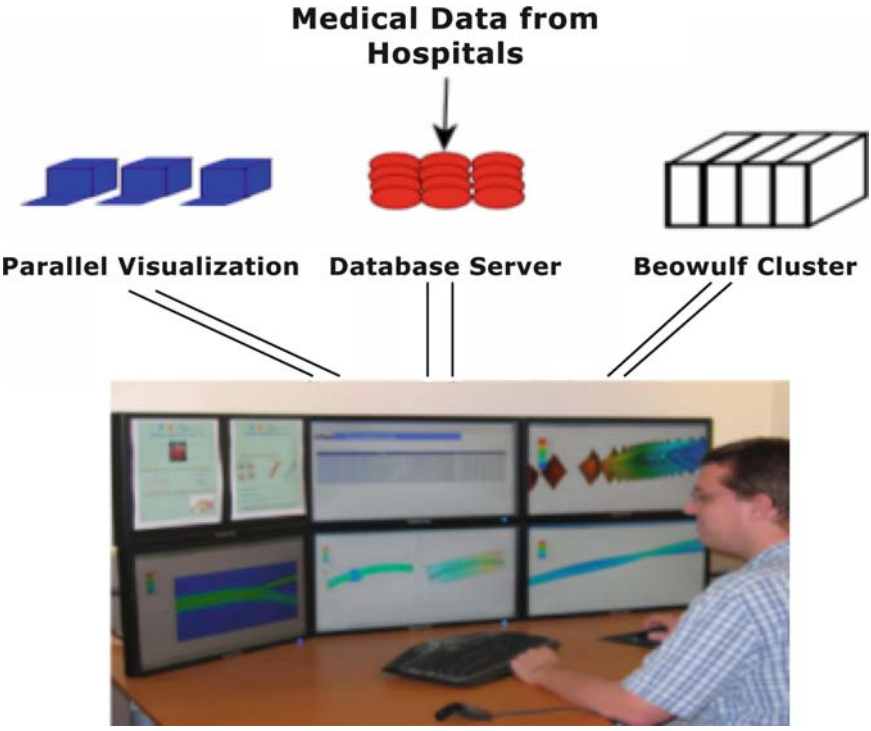


Fig. 2 Second generation: IDV desk

Fig. 3 Third generation: CDS system

these has a high-end graphic card with two DVI outputs and controls a column of two displays. One of the drawbacks of this assembly is the heat generated by the graphic cards and the multicore processors. In order to limit the noise that would be generated by a conventional fan, we use water cooling to remove the heat from the workstations. Overall, the CDS system hardware is fairly silent and can be installed in a standard office.

The fourth workstation is a data server with two sets of disks. It has two slow 7,200 rpm disks with large capacities for storage (1 TB) and one fast 10,000 rpm disk with less storage (150 GB). The goal of the fast disk is to prefetch the data to speed up data mining. This amount of storage is adequate for 1,500 patient studies.

The overall cost of the CDS system is less than $12K for 16 cores and 16 GB of main memory. From the hardware point of view, it is a flexible system where each component can be customized to some degree. To make this system effective, we have to combine several interesting open source software. We will describe in the next section the various software components of the system.

The idea behind the CDS system is to allow the simulation and the visualization of multidimensional data. The system is flexible on the software side as much as it is on the hardware side. The goals are to compute, store, visualize, and retrieve data in a secure way. Figure 4 shows the integration of all these subsystems.

As for the hardware, each component is chosen to perform a specific task. The underlying OS running on the three workstations in charge of the visualization and the computation is Linux. It allows to have a system that is customizable, secure and it is one of the most used operating systems for running parallel computations. It is also more stable for 64 bits architecture than some other OS. The database computer runs also Linux. For the ease of development of the database architecture, and the tight integration with its environment, the choice was made to develop the database with mySQL.

The other components that we integrate in the CDS system are off-the-shelf softwares distributed with open source licenses. For instance, to avoid to be encumbered by several mouse and keyboards for each workstation, we can control the CDS system with a single keyboard and a single mouse thanks to the software Synergy [40].

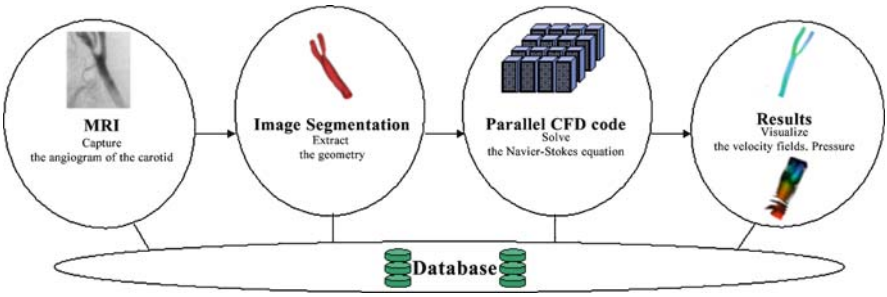


Fig. 4 Hemodynamic simulator linked to the database

Synergy allows sharing a single mouse and keyboard between multiple computers with different operating systems, each with its own display, without special hardware. It is intended for users with multiple computers on their desk since each system uses its own monitor(s). Redirecting the mouse and keyboard is as simple as moving the mouse off the edge of the screen. Synergy also merges the clipboards of all the workstations into one, allowing cut-and-paste between them.

Another way of controlling the information that is displayed on the CDS system is to use a laptop or a tablet PC. This allows the control of the important data remotely. In order to perform this task, we need to achieve remote desktop control. Using the software such as rdesktop/Krdc, Krfb and Kompose, we can easily manage the information graphically from any remote device. Kompose offers a great new way to manage a big number of windows on different virtual desktops. It creates a full screen view where every window is represented by a scaled screenshot of itself.

The CDS system might be used not only for surgery planning by building surgery scenarios in order to evaluate risk but also for postsurgery follow up and evaluation of the improvements. Our initial case study is about carotid surgery intervention in the neck. Identifying stenose, plaque location, processing detailed hemodynamic simulation at the bifurcation are at the core of this process. Additional application domains for the CDS system would be for example:

- Managing endovascular graft: deformation/displacement of stent, hemodynamic in low extremity vessel, effect of physical activity, etc.
- Brain aneurisms: identifying low shear stress area, high pressure zone, area where the wall is too weak, anticipating the impact of the placement of coil, etc.
- Breast conserving therapy and tumor treatment planning: from tissue mechanic and tumor growth dynamic, planning tumor removal and radiotherapy with minimum invasive procedures, etc.

As said earlier, we will focus here on the treatment of the carotid complex on the neck. The carotids supply the brain with freshly oxygenated blood. Blockage of these arteries occurs over time through atherosclerosis or build up of plaque most often at the bifurcation location where internal and external branches start.

Small pieces of the plaque called emboli can break off and occlude branch arteries in the brain or the eyes. Shortage in blood supply may give temporal loss of brain function (=ischemic event) or permanent damage (=cell dies): those are strokes. Two main classes of surgery intervention are endarterectomy and/or repair. Carotid endarterectomy is a surgical operation that removes a flatty plaque. With endovascular repair, a minimum invasive surgical operation open up the lumen at the stenose location by placing a stent. However, this operation is subject to restenosis. The decision process to go for one type of intervention to the other might be enhanced by a systematic study that combines multiple medical image interpretation, including hemodynamic simulation and patient background analysis.

Figure 5 shows the typical workflow for the CDS system. In the subsections, we will describe each process step (I–VII) and the data input and output of each step.

Now we are going to describe each component of the system design following the workflow that starts from a new patient study.

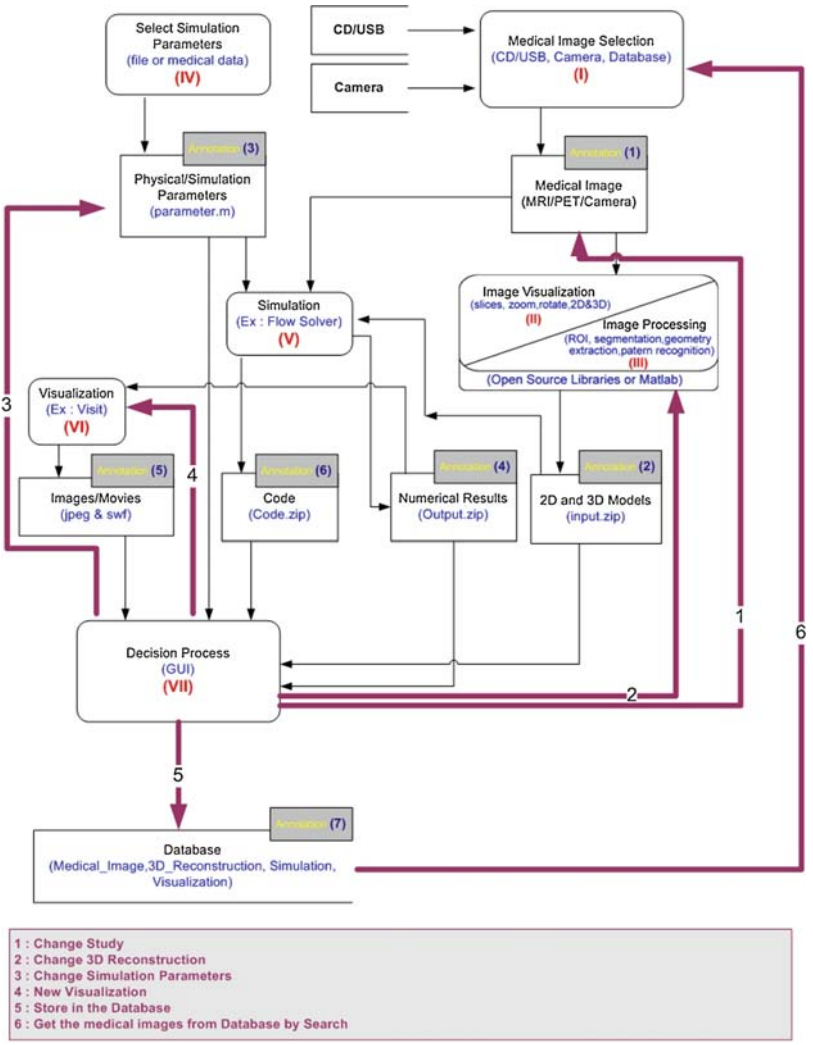


Fig. 5 Computational desk for surgeons (CDS) workflow

Step I: Medical Image Selection

The medical image that is the input to the CDS system can come from several sources. One is a new medical image for a new patient. It comes on a CD, DVD, or USB memory stick. The second source can be the database.

The challenge is to provide a database organization and interface that can be flexible enough to integrate all modalities involved in image analysis and image based simulation in order to evolve with the constant progress made in this application domain. Further, the interface should be intuitive enough to be used by noncomputer

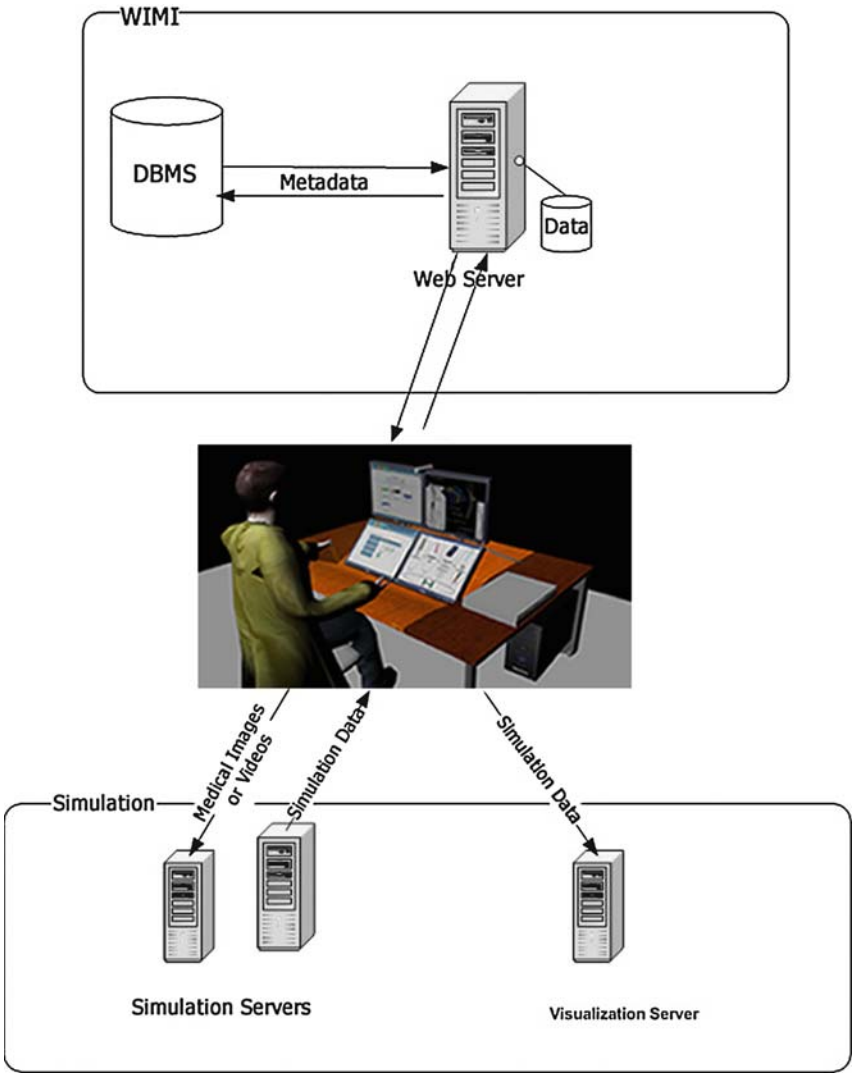


Fig. 6 Computational desk for surgeons (CDS) architecture

specialists. In Fig. 6, a high level representation of the system architecture is given. This model was chosen because it offers the most flexibility to the end user. Access to CDS is limited to those that are registered users with valid passwords. The passwords are encrypted and stored in the database itself. CDS system is designed with a novel idea of integrating the simulation datasets with medical images which needs further flexibility. We originally planned to use the popular DICOM schema as it is, but storing the simulations and images together needed some modifications of the DICOM schema.

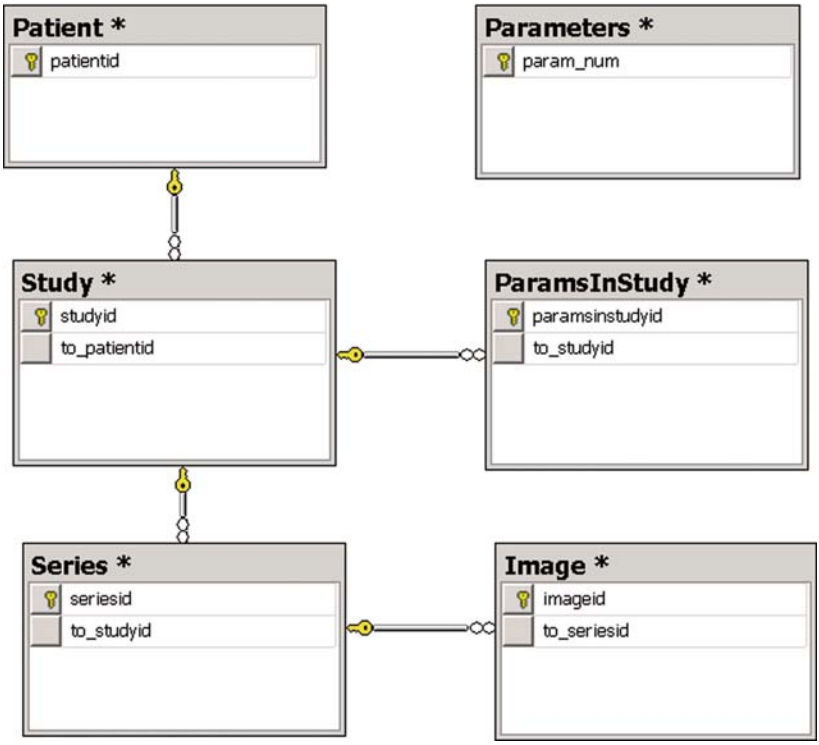


Fig. 7 Computational desk for surgeons (CDS) database schema

The DICOM database schema contains tables for patient, study, series, and images. Using this schema, all the images created for many series of studies conducted on a patient can be stored. CDS system also does the same, and in addition stores simulations created using the medical images. In order to store this, we added two more tables: Parameters and ParamsInStudy (Fig. 7).

CDS system offers three different searching techniques: search by *comments,* search by *parameters,* and search by *patients* through which the users can retrieve the required information. The nature of the data stored in the database creates the need for search based on different factors. The database schema used in CDS system makes these searches efficient.

Step II: Image Visualization

CDS system uses software, for example ImageJ, to allow the doctor to visualize the original medical image. The surgeon can zoom in/out, rotate, get 2D surface rendering, and 3D volumetric rendering of the medical image. Example of rendering are in Figs. 8–10.

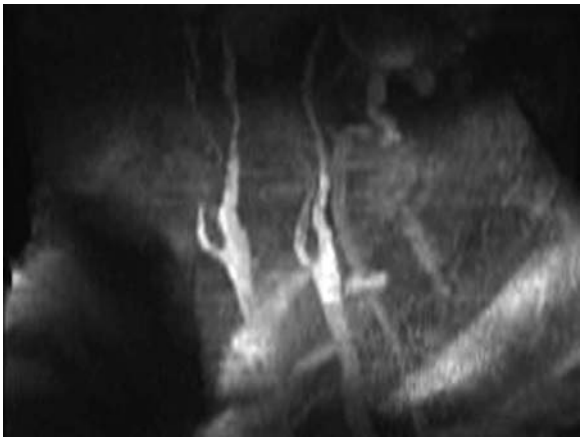


Fig. 8 MRI: frontal projection (left and right is swapped)

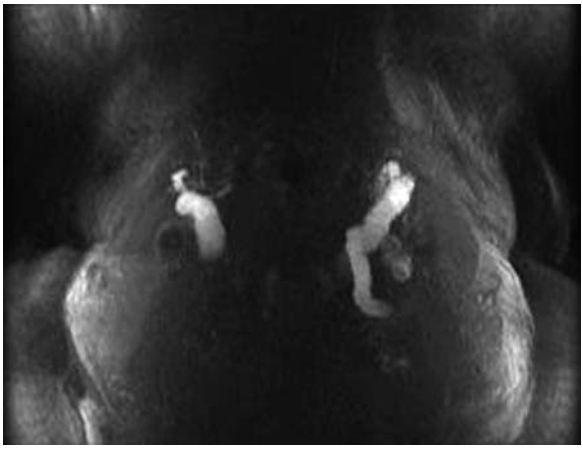


Fig. 9 MRI: from the top projection

Step III: Image Processing: Modeling

The CDS system displays projected images to facilitate in the pattern recognition process, i.e., the doctor understands from his/her experience where is the ROI. See Figs. 11 and 12.

The simulation must use an accurate geometry of the carotid that can be obtained from the MRI slices. In these images, blood, i.e., the artery lumen is bright, while tissue is dark. We know a priori the image dimension. In one of our example, dimensions for the axial images of each set are $0.625\,\mathrm{m} \times 0.625\,\mathrm{m}$. The thickness of each image is 2 mm. Each image is a 3D slice with an in-plane resolution of

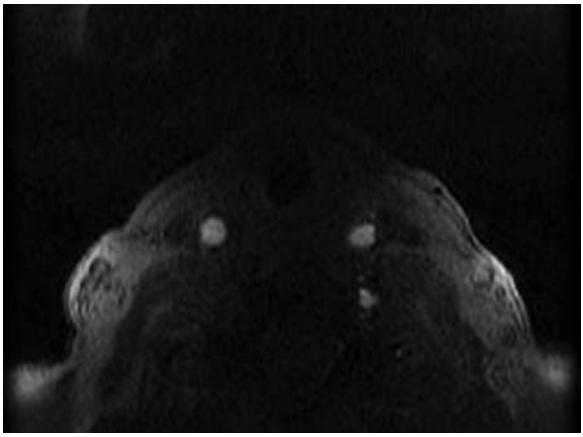


Fig. 10 MRI slices reconstruction

Fig. 11 ROI for the frontal projection (left and right is swapped)

0.625 mm and a slice thickness of 2 mm. We may get several hundred slices or much less depending on the resolution of the MRI. One selects on the touchscreen a square on both projection image that contains the ROI. The system can then compute the set of MRI slices and define the square subdomain on each slice that will be used. Eventually, one can refine the ROI from this new set of slices.

Using a simple threshold algorithm, each slice might be transformed into a black and white figure that provides the corresponding lumen slice. Image segmentation and noise reduction might be more reliable when it is done interactively by the surgeons. We will also include a broad spectrum of standard image segmentation tools including some level set algorithms. Using our own code, we can reconstruct a 3D

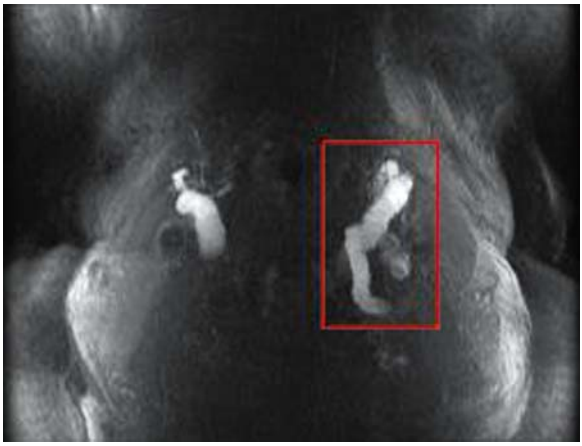


Fig. 12 ROI for the top projection

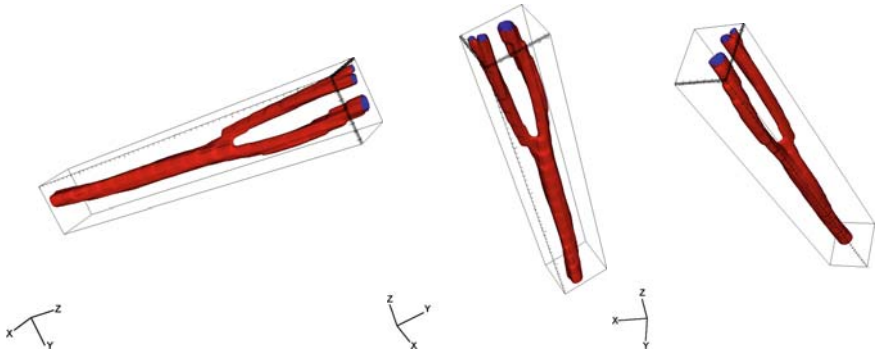


Fig. 13 Construct segmental image

spatial structure that represents the lumen for the section of carotid. This reconstruction requires a much finer grid resolution that the MRI data. We use interpolation and postsmoothing. An example of 3D reconstruction of the segmented image is shown in Fig. 13.

Step IV: Select Simulation Parameters

From this 3D reconstruction, one can abstract a representation that can be used for a posteriori statistical purposes. For example, one extract the skeleton of the carotid bifurcation (central line), and provide measurement of two to three characteristic angles at the bifurcations. At each section of the artery branches, one extract inner and

outer diameter of fitted rings. The ratio of those diameters provided an indicator of asymmetry. Along these lines, a coarse geometric model may quantify the stenose, give indicator on the plaque location, be used in a coarse flow analytical model based on a Wormesley solution or some even simpler model.

Assuming inflow/outflow boundary conditions, we may get instantaneously coarse blood flow analytical approximation, extremum velocity and debit indicators. We cannot claim that these analytical approximations are very accurate, but they are rather indications that may have statistical values with large data sets.

Accurate information on the hemodynamic flow should be derived thanks to a direct simulation of the Navier–Stokes equations. Because we deal with relatively large arteries, the simulation is done with an incompressible Navier–Stokes (NS) model. While such a NS code requires a large number of set up parameters, for a class of problem, such as the carotid complex in the neck, many of those can be set up once for all and hidden from the user. We need however the user to select the input wave signal for the velocity flow or inlet pressure, and the type of output boundary conditions. Typically, one has to specify the distribution of the flux between the two branches. The periodic wave form of the inflow should be taken either from a file that contains a measurement: in this case, it is a one dimensional array that represents a time period. Alternatively, one can preset a wave signal with fewer parameters that belong to a category of interest.

The inflow parameters can be obtained either from the general pressure wave form, see Fig. 14. In such a case, we have to give the maximum pressure (systolic), minimum pressure (diastolic) and period (heart bit pulse). The wave form can be

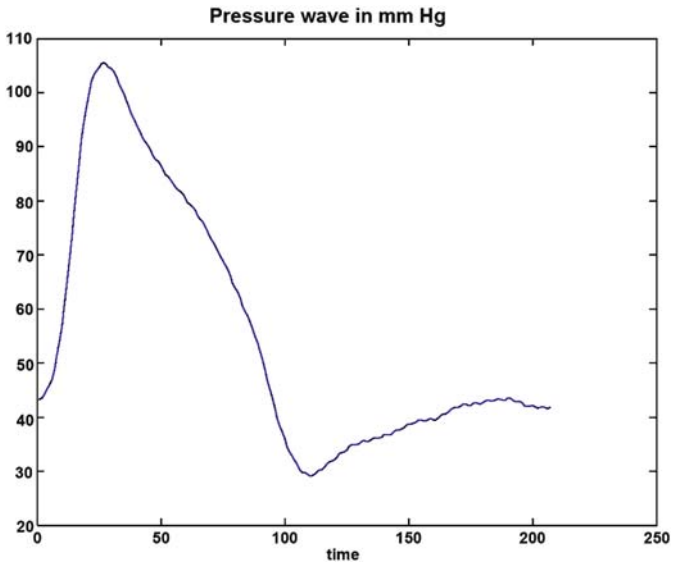


Fig. 14 Inflow parameters: general pressure wave form

taken from few representative classes of wave forms characteristic of the health condition (heart decease, hypertension, etc.).

Or, the inflow parameters can be obtained from an ultrasound image that provides information on blood velocity input. One can possibly use phase contrast MRI data, that specifies the inflow profile at a given location (downstream velocity component for example) as well as the outflow boundary conditions.

We define $\theta \in (0, 1)$ to be the distribution of flow between the two branches. θ must be chosen by the user, and might be derived from medical imaging.

Eventually, the bifurcation might be coupled to a down stream network model. This is key for surgery procedures that will definitively change the outflow boundary conditions, but may not be relevant for carotid bifurcation in the neck.

Step V: Simulation

The numerical simulation is based on the incompressible Navier–Stokes equations to model the flow, inlet and outlet blood flow conditions, and a level set method to derive the image segmentation of the artery in the ROI. Level set image and Navier–Stokes flow are coupled thanks to an immersed boundary method that is first order accurate. However, the method is numerically robust. This technique avoids the need for meshing of three dimensional complex objects that is often the bottleneck in finite element simulation. Our method is fairly simple since we can work with Cartesian grid and is very efficient on parallel architectures with standard scalars processor thanks to a domain decomposition technique. The interested reader can find more details on our flow solver and domain decomposition technique in [41].

Rigorous error estimate with unsteady Navier–Stokes flow in the context of finite volume approximation is not well covered by the current applied mathematic literature. We have however used the classical concept of manufactured solution and our optimized extrapolation method to build error indicators in the simulation environment [42]. This solution is far from perfect but may still be useful to the end user.

Navier–Stokes Equations

The hemodynamic properties of the blood flow can be approximated by the incompressible Navier–Stokes equations for the large arteries [43].

For fast prototyping of incompressible NS flow, the Eulerian approach is considered and one can combine fast solvers for regular Cartesian grid solution with some form of fictitious domain decomposition or locally fitted stencil to implement the boundary conditions [44, 45]. The penalty method, introduced by Caltagirone and coworkers [46], is used here since it is simpler to implement and applies naturally to flow in a pipe with moving walls [47].

For simplicity, we describe the method in the two dimensional space case. The flow of incompressible fluid in a rectangular domain $\Omega = (0, L_x) \times (0, L_y)$ with prescribed values of the velocity on $\partial\Omega$ obeys the NS equations:

$$\partial_t U + (U \cdot \nabla)U + \nabla p - \nu \nabla \cdot (\nabla U) = f, \text{ in } \Omega$$

$$\text{div}(U) = 0, \text{ in } \Omega.$$

We denote by $U(x, y, t)$ the velocity with components (u_1, u_2) and by $p(x, y, t)$ the normalized pressure of the fluid. ν is a kinematic viscosity.

With an immersed boundary approach, the domain Ω is decomposed into a fluid subdomain Ω_f and a wall subdomain Ω_w. In the L_2 penalty method, the right hand side f is a forcing term that contains a mask function Λ_{Ω_w}

$$\Lambda_{\Omega_w}(x, y) = 1, \text{ if } (x, y) \in \Omega_w, \ 0 \text{ elsewhere},$$

and is defined as follows

$$f = -\frac{1}{\eta} \Lambda_{\Omega_w} \{U - U_w(t)\}.$$

U_w is the velocity of the moving wall and η is a small positive parameter that tends to zero.

A formal asymptotic analysis helps us understand how the penalty method matches the no slip boundary condition on the interface $S_w^f = \bar{\Omega}_f \bigcap \bar{\Omega}_w$ as $\eta \to 0$. Let us define the following expansion:

$$U = U_0 + \eta U_1, \ p = p_0 + \eta p_1.$$

Formally we obtained at leading order,

$$\frac{1}{\eta} \Lambda_{\Omega_w} \{U_0 - U_w(t)\} = 0,$$

that is

$$U_0 = U_w, \text{ for } (x, y) \in \Omega_w.$$

The leading order terms U_0 and p_0 in the fluid domain Ω_f satisfy the standard set of NS equations:

$$\partial_t U_0 + (U_0 \cdot \nabla)U_0 + \nabla p_0 - \nu \nabla \cdot (\nabla U_0) = 0, \text{ in } \Omega_f$$

$$\text{div}(U_0) = 0, \text{ in } \Omega.$$

Further, the wall motion U_w must be divergence free. It was shown in [48] for fixed wall, i.e., $U_w \equiv 0$, that the convergence order of the penalty method is of order $\eta^{\frac{3}{4}}$, in the fluid domain, and $\eta^{\frac{1}{4}}$ in the wall.

To acquire the geometry of a large vessel from medical images, the level set method based on the Mumford–Shah energy model [49], has been implemented. This combines perfectly with the penalty method which allows us to deal with complex geometries such as carotid bifurcation or the circle of Willis.

Integration of the Image Analysis

Doctors and medical surgeons use different technologies such as ultrasound, computer tomography (CT) or magnetic resonance imaging (MRI) in order to take images of the vessels. CT or MR angiography images are used to extract the exact geometry of the arteries that are specific to each individual. This process is crucial to ensure the accuracy in computational fluid dynamic (CFD) computations.

The mask function Λ_{Ω_w} of the penalty method, is obtained *directly* with an image segmentation technique that is a level set method. Since the contours of the image are not necessarily sharp, it is interesting to use the level set method presented in [49] and based on the Mumford–Shah Model. For completeness, we are going to briefly describe this method. We refer to the review papers [50, 51] for a more comprehensive description of the level set method in the framework of image analysis. Let us denote $C(s)$ the unknown parameterized curve(s) that delineate the vessel. We assume that the unknown function(s) $C(s) : [0, 1] \to IR^2$ is a piecewise $C^1[0, 1]$ function.

In the level set method, $C(s)$ is represented by the zero level set of a Lipschitz function: $\phi : \Omega \to IR$.

$C(s)$ should correspond to the minimum of the energy $F(C, c_1, c_2)$:

$$F(C, c_1, c_2) = \mu_1 \cdot (\text{length}(C)) + \mu_2 \cdot (\text{area}(\text{inside}(C))) + \lambda_1 \int_{\text{inside}(C)} |u_o - c_1|^2 \, \mathrm{d}x + \lambda_2 \int_{\text{outside}(C)} |u_o - c_2|^2 \, \mathrm{d}x,$$

where

$$c_1(\phi) = \frac{\int_\Omega u_0 \, H(\phi) \, \mathrm{d}x}{\int_\Omega H(\phi) \, \mathrm{d}x},$$

$$c_2(\phi) = \frac{\int_\Omega u_0 \, (1 - H(\phi)) \, \mathrm{d}x}{\int_\Omega (1 - H(\phi)) \, \mathrm{d}x}.$$

H is the Heaviside function $H(z) = 1$, if $z \geq 0$, 0 if $z < 0$. To understand the energy function used in this model, let us suppose that μ_1 and μ_2 are set to zero, and let us suppose that the image is a piecewise constant function with values 0 and 1. The angiogram of an artery with a 50% stenosis given in Fig. 15 is at first sight close to this description. Clearly, the functions $C(s)$ that realize the minimum of energy are the boundaries that delimit the two sets $u_0 = 0$ and $u_0 = 1$. The first two terms in the energy model are common in many active contour methods and control the smoothness of the curve $C(s)$ as well as the detection of the edges.

Fig. 15 Angiogram

The numerical process to compute ϕ uses the following evolution problem:

$$\frac{\partial I}{\partial t} = N[I],$$

where N is the associated Euler Lagrange equation

$$N[\phi] = \frac{\mathrm{d}}{\mathrm{d}\, z} H(z) \left[\mu \ \mathrm{div}\left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \nu - \lambda_1 (I - c_1)^2 + \lambda_2 (I - c_2)^2 \right].$$

In the numerical implementation, one uses a regularization of the Heaviside function $H_\varepsilon \in C^1(0,1)$ such that $H_\varepsilon \to H$, as $\varepsilon \to 0$, as well as a reinitialization procedure every few time steps that sharpens the level set function in the neighborhood of the zero level set.

To successfully apply the Chan–Vese method, one must carefully choose the parameters of the method, i.e., $\lambda_1,\ \lambda_2,\ \mu_1,\ \mu_2$. From our experience, a successful combination of parameters to detect the edge of a thin long object is in general such that with $\lambda_1 >> 1$, and $\lambda_2 = 1$. This follows somehow the fact that the area inside $C(s)$ is much smaller than the area outside $C(s)$.

The mask function Λ_{Ω_w} in the momentum equation is consequently

$$\Lambda(x, y) = H(\Phi(x, y)), \quad (x, y) \in \Omega.$$

There is an obvious advantage in combining the level set method with the penalty technique for NS because the level set function provides directly the source term in the momentum equation. We do not need an explicit geometric representation of the wall boundary as a differentiable deformable model will do. Further, any artifact

Fig. 16 Image segmentation

possibly produced by the level set method such as an artificial closed subset of fluid in the wall, such as in Fig. 15, should have little impact on the computation of the flow in the artery. As a matter of fact, such cavity has $U \approx U_s$ on all boundaries.

Figure 15 gives some representative example of a two dimensional section of an angiogram, and Fig. 16 a corresponding image segmentation.

Thank to the image segmentation, we can simulate the blood flow through a real artery. Since we are looking for faster result, we need to have an efficient solver for the resolution of the NS equations. We use a domain decomposition algorithm called Aitken Schwarz that has high numerical efficiency and scales well on low cost Beowulf Clusters [52, 53]. This is also advantageous for our low cost CDS system that uses a standard Gigabyte Ethernet switch to connect the fat nodes.

Fast Solver

To solve the incompressible Navier–Stokes equations, we use a projection method [54] for the time step as follows (we denote u^k, respectively p^k the numerical approximation of velocity U, respectively pressure p at time step k):

- Step 1: prediction of the velocity $\hat{u}^{k+1}$ by solving

$$\frac{\hat{u}^{k+1} - u^{k,*}}{\Delta t} - \nu \Delta u^k = f^{k+1} - \nabla p^k$$

in $(0, L_x) \times (0, L_y)$ with the boundary condition $\hat{u}^{k+1} = g$ on $\partial\Omega$. The ratio $\frac{\hat{u}^{k+1} - u^{k,*}}{\Delta t}$ is a first order approximation of the total derivative in time and $u^{k,*}$ comes from the methods of characteristics. Eventually, the diffusion term in this equation can be made implicit depending on the local ratio between ν and the space mesh size.

- Step 2: projection of the predicted velocity to the space of divergence free functions.

$$-\text{div}\nabla\delta p = -\frac{1}{\Delta t}\text{div}\hat{u}^{k+1}, u^{k+1} = \hat{u}^{k+1} - \Delta t\delta p,$$

$$p^{k+1} = p^k + \delta p.$$

This set of equations is discretized with a standard finite different scheme on staggered grids.

In an incompressible Navier–Stokes code, the resolution of the pressure correction, that corresponds to a Poisson problem, is the most time consuming routine. Indeed, depending on the size of the problem, between 50% up to 97% of the elapsed time of the whole code is dedicated to the resolution of this linear system [55]. This is exemplified in Fig. 17; for 16 millions unknowns, the resolution of the pressure solver for our incompressible Navier–Stokes 3D simulation represents almost 90% of the elapsed time of the sequential processing. It is then critical to have a fast parallel solver for the pressure correction.

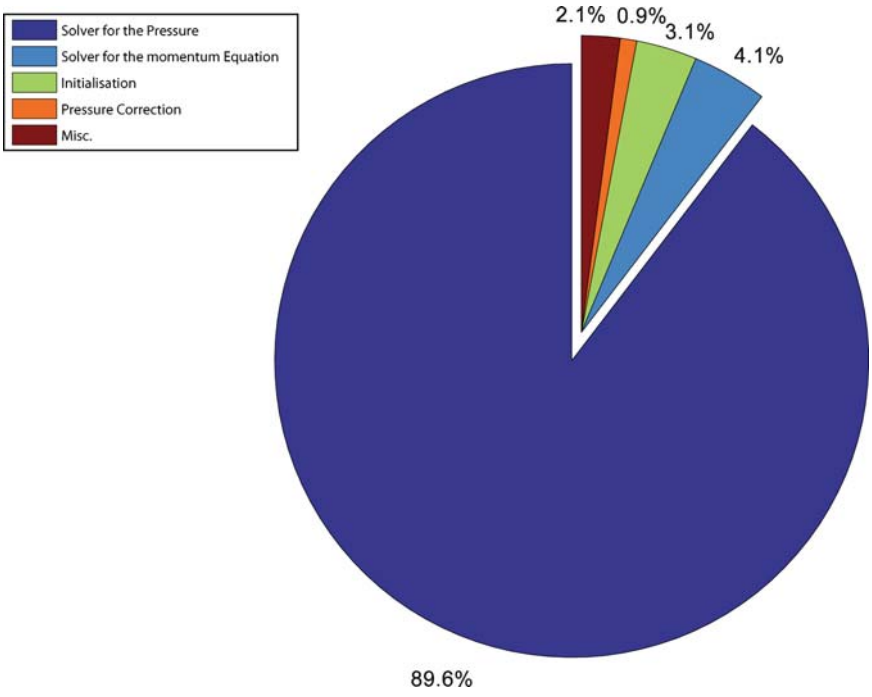


Fig. 17 Pie representing the percentage of the elapsed time routines on an incompressible Navier–Stokes 3D code with 16 millions of unknowns

Solver for the Pressure Equation

The pressure equation can be integrated with a number of existing fast Poisson solvers since the discretization grid is regular. The parallel version of our solver uses the Aitken–Schwarz domain decomposition (DD) technique presented in [53, 56]. This DD technique is interesting for distributed computing with high latency network since it provides good scalability results [57]. For each subdomain, we choose the subdomain solver for the Poisson problems that provides the smallest elapse time. We refer to [58] for an extensive study of the performance comparing several libraries invoking either an LU decomposition, or a Krylov method, or a multigrid scheme. We will see later on that our NS solver can perform very quickly the numerical simulation. However, the main difficulty with the penalty method consist in retrieving accurately the shear stress on the wall boundary. As a matter of fact, the rhs of the momentum equation exhibits a stiff jump at the wall that forces the velocity to match the no slip boundary conditions but leads also to numerical instabilities in the velocity profile through the wall. We are going to address this problem in the next section.

Computation of the Shear Stress

While the computation of the hydrodynamic forces exerted by the fluid on the wall is easy to compute using the integral on Ω_w of the penalty term [47], the computation of the shear stress is more problematic in the penalty method. We restrict ourselves to a two dimensional description for simplicity, while the computation has been extended successfully to three space dimensions.

The shear stress in the boundary layer can be obtained from the formula

$$\tau = \nu \left(\frac{\partial w_1}{\partial \xi} + \frac{\partial w_2}{\partial \eta} \right),$$

where (ξ, η) is the normal/tangential coordinate system along the wall, and (w_1, w_2) are the components of the flow field along respectively the tangential and normal direction to the wall. Because the flow field (u_1, u_2) is computed on the Cartesian staggered grid, we rewrite the shear stress formula in the (x, y) coordinate system. If α denotes the angle of the tangent to the wall at point $M(x, y) \in S_w^f$, with the horizontal axis x, we get

$$\tau = \nu \left(\cos(2\alpha) \left(\frac{\partial u_1}{\partial y} + \frac{\partial u_2}{\partial x} \right) + \sin(2\alpha) \left(\frac{\partial u_2}{\partial y} - \frac{\partial u_1}{\partial x} \right) \right).$$

The flow field is continuous, but not differentiable on S_w^f . We cannot, therefore, approximate the shear stress with some central finite differences formula that will require points on both sides of the wall. Further, the computed velocity exhibits small oscillations in the vicinity of S_w^f inside Ω_f because of the stiffness of the forcing term in the momentum equation at that location.

We use a gridless approach for the shear stress computation. Let R be the rectangle $(-d, d) \times (0, l)$ in the (ξ, η) coordinate system that is tangent to the wall at $M(x, y) \in S_w^f$ and lies inside the flow region. Let R_h be the set of grid points inside this rectangle, completed by the grid points obtained from the intersection of the Cartesian mesh with the wall in the ball of center $M(x, y)$ and radius d. For each flow velocity components one computes a second order polynomial approximation $P_M(x, y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy$ that fits in the least square sense the flow field component on R_h. We notice that the least square approximation filters out the possible oscillation of the solution near the wall.

The derivatives of the velocity field are then approximated by the derivative of $P(x, y)$. The dimensions of the rectangle are as follows. The width of the rectangle R is chosen to include approximatively three grid points, i.e., $d = \sqrt{h_x^2 + h_y^2}$. The length l of the rectangle is chosen to be proportional to the boundary layer thickness, i.e., $l \approx \sqrt{\nu}$, for small ν, in order to capture accurately the parabolic profile of the flow field in the layer. l is a priori independent of the mesh size.

We have carefully checked the sensitivity of the accuracy of the computation of the shear stress, for a simple Poiseuille flow in a straight pipe, by translating and rotating the pipe in the fixed Cartesian grid. Our experiment suggest that we can get an error of less than 5% with grids of moderate size. This level of accuracy is perfectly acceptable for image based CFD of blood flow in main arteries. The verification of the solution of our code was done also by comparing our simulation with the solution computed by ADINA for Reynolds number less than $1,000$ [59]. Similar results were obtained in Bilel Hadri's PhD dissertation [60] for the three dimensional case.

Now we are going to describe the parallel performance of our image base simulator.

Performance Evaluation

In this section, we provide the performance of our 3D parallel code. A large number of geometries corresponding to real medical test cases as well as artificial benchmark problems have been run for code verification purpose.

Figure 18 represents the elapsed time in seconds for one time step depending on the number of subdomains and the grid size for a sequential code. Because we use a direct solver for the pressure, the elapse time per time step is dependent on the discretization mesh size only. Consequently, our results can be reliably used to predict the elapse time of a simulation.

We observe that we obtain the best performance with an optimal number of subdomains that grows with the size of the discretization mesh. In general, the parallel algorithm, running on a single processor, is (much) faster than the sequential algorithm, that is with no domain decomposition. It appears that the Aitken–Schwarz algorithm is Cache friendly [53].

In Table 1, we report the acceleration of the method depending on the number of subdomains and the grid size. For some configurations, we get an acceleration of

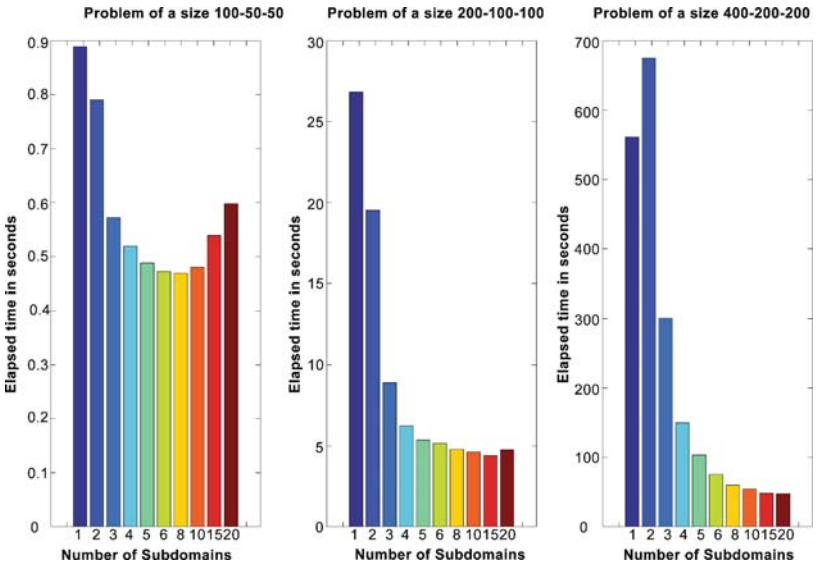


Fig. 18 Elapsed time in second for one step time depending on the number of subdomains and the grid size

Table 1 Acceleration of the method depending on the number of subdomains and the grid size

	Number of subdomains							
Grid size	2	4	5	6	8	10	15	20
0.25×10^6	1.1	1.7	1.8	1.9	1.9	1.8	1.6	1.5
2×10^6	1.4	4.3	5.1	5.2	5.6	5.8	6.1	5.6
16×10^6	0.8	3.7	5.4	7.4	9.4	10.3	11.7	11.8

the method close to 12 times. We can notice that if we multiply the grid size of the problem by 8, the elapsed time is multiplied by 10. Indeed, for 250,000 unknowns, we spent 0.5 s per time step, for 2,000,000 unknowns around 5 s per time step and 16,000,000 unknowns close to 50 s per time step. We have a quasi linear behavior of the elapsed time when we increased the size of the problem.

In Table 2, we present the parallel version running on a 24 SUN X2100 nodes, 2.2 GHz dual core AMD Opteron processor, 2 GB main memory each, with an Infiniband Interconnect. The table shows the speedup based on the sequential code on a single processor reached 95% parallel execution. Overall, one expects a carotid simulation to take about 1 min per cardiac cycle with the CDS system.

Step VI: Visualization

The visualization is done using Visit (http://www.llnl.gov/visit/) which is developed by Lawrence Livermore National Laboratory. This software allows to display

Table 2 Elapsed time for the resolution of one step time

	Number of processors							
Grid size	2	3	4	6	8	16	20	24
100–50–50	0.52	0.1	0.24	0.17	0.13	0.2		
200–50–50	2.08	0.98	0.58	0.34	0.26	0.23	0.29	
400–50–50	12.80	3.72	2.13	1.03	0.62	0.36	0.36	0.41
300–75–75	10.39	4.96	2.89	1.53	0.93	0.67	0.69	
200–100–100	8.91	4.21	2.46	1.4	1.02	0.89	1.15	1.58
400–100–100	34.8	15.5	9.08	4.34	2.68	1.48	1.41	
400–200–200	188	67.7	40.2	20.16	12.65	6.65	6.32	

multidimensional data and to navigate synchronously in space and time between different modalities. Visit also offers the possibility to have predefined behavior through its scripted engine. This, for example, allows a surgeon to run a simulation and automatically have access to the resulting set of data. On the performance side, Visit has native parallelism in its core engine and therefore, can make use of multicore/multiprocessors systems.

Step VII: Decision Process

At this point in the workflow, the surgeon can decide to redo any of the previous steps (see Fig. 5), such as:

1. Change the study
2. Change the 3D reconstruction
3. Change simulation parameters
4. New visualization
5. Store the data in the database

This loop should be run as many times as necessary and will provide huge archives. It is only from long term clinical experience that one can learn which steps in this process will be critical for clinicians and which ones will have marginal impact.

Conclusion

We have presented a cost effective solution that integrates simulation, visualization, and data management into a stand alone digital processing desk. Our prototype system has been designed to efficiently process the workflow in planning and/or studying endovascular surgery. Many of the concepts and solutions presented can be used in other applications, where the decision process relies on the efficient

and interactive (parallel) processing and data management of images, modeling, and simulations. We feel however that it is critical to close the interaction loop between the surgeon and the system. Many technical refinements might be added such as adding voice navigation through the application or the surgeon biometric recognition to our CDS system. But most importantly is the concept of a multi-disciplinary team working with surgeons to bring the CDS system at the level of reliability and efficiency requested by clinical practitioners. In such collaboration, the product should be the design of new methods and tools to solve one of the most common and difficult problems in computational surgery, i.e., really use computers for measurable clinical care improvement.

References

1. Greenlund KJ, Giles WH, Keenan NL (2006) Heart disease and stroke mortality in the 20th century. In: Ward J, Warren C (eds) Silent victories: the history and practice of public health in twentieth century America. Oxford University Press, Oxford, England
2. Thom T, Haase N, Rosamond W, Howard VJ, Rumsfeld J, Manolio T et al (2006) Heart disease and stroke statistics – 2006 update. Circulation 113(6):e85–e151
3. Garbey M, Picard C, Hilford V (2008) Toward an intelligent data and visualization desk for endovascular surgery. In: Conference of computers and their applications CATA 2008, Cancun, Mexico
4. Hilford V, Garbey M, Vadakattu S, Addison C (2007) WIMI – Web Interface to Medical Information, best paper award in data engineering at SEDE'07, 16th internat. In: Conference on software engineering and data engineering, July 2007, Las Vegas, USA
5. Maisel JM, Wisnicki HJ (2002) Documenting the medical encounter with speech recognition: physicians can adopt this new technology to increase the turnaround time of quality records. Ophthalmology Times 27(5):38–41
6. Gosbee J, Ritchie E (1997) Human–computer interaction and medical software development. Interactions 4(4):13–18
7. Zick RG, Olsen J (2001) Voice recogition software versus a traditional transcription Service for physician charting in the ED. Am J Emerg Med 19(4):295–298
8. Roop ES (2006) Speaking of savings: can speech recognition deliver? For the Record 18(10):24
9. Srinivasan S, Haque S, Mital D Java enterprise wide application for clinical databases. In: 7th international conference on control, automation, robotics and vision (ICARCV'OZ), DK 2002, Singapore
10. Zudilova EV, Sloot PMA, Belleman RG. A multi-modal interface for an interactive simulated vascular reconstruction system. In: Proceedings of the fourth IEEE international conference on multimodal interfaces. Section of Computational Science Faculty of Science, University of Amsterdam
11. Wang F, Rabsch C, Liu P (2007) Using SVG to model and query image annotations and their history. IEEE Int Conf Bioinform Biomed 412–422
12. Stanford University. iPAD, Stanford University. Available at: http://bimm.stanford.edu/main/ipad
13. Northwestern University, Stanford University. Annotation Imaging Markup (AIM), Northwestern University, Stanford University. Available at: https://cabig.nci.nih.gov/tools/AIM
14. Philippe PUECH, Loc BOUSSEL. DicomWorks. Available at: http://dicom.online.fr/
15. DesAcc Inc. Digital Jacket Pro, DesAcc Inc. Available at: http://www.desacc.com/products/djpro/index.html
16. TeraRecon Inc. TeraRecon, TeraRecon Inc., San Mateo, CA. Available at: http://www.terarecon.com/index.php

17. CoActiv Medical Business Solutions. CoActiv, CoActiv Medical Business Solutions, Ridgefield, CT. Available at: http://www.coactiv.com/
18. Biotronics3D Ltd. Biotronics3D, Biotronics3D Ltd, London, England. Available at: http://www.biotronics3d.com/
19. Guthrie B. Emageon, Division of Neurosurgery at University of Alabama in Birmingham (UAB). Available at: http://www.emageon.com/
20. Rasband W. ImageJ, National Institutes of Health, Bethesda, MD. Available at: http://rsb.info.nih.gov/ij/
21. Lee H-W, Liu B-D, Hung K-C, Lei S-F, Wang P-C, Yang T-L (2009) Breast tumor classification of ultrasound images using wavelet-based channel energy and imageJ. IEEE J Sel Top Signal Process 3(1) 81–93
22. Cannataro M, Guzzi PH, Tradigo G, Veltri P. A tool for the semiautomatic acquisition of the morphological data of blood vessel networks. In: International symposium on parallel and distributed processing with applications, Laboratory of Bioinformatics, University Magna Grcia of Catanzaro, Catanzaro, Italy, 2008
23. Sage D, Unser M (2003) Teaching image-processing programming in Java. IEEE Signal Process Mag 20(6):43–52
24. Agrafiotis D, Bull DR, Canagarajah N (2003) ROI coding of volumetric medical images with application to visualisation. In: Proceedings of the 3rd international symposium on image and signal processing and analysis, Image Communications Group, Centre for Communications Research, University of Bristol, UK
25. Yang Y, George S, Martin DR, Tannenbaum AR, Giddens DP. 3D modeling of patient-specific geometries of portal veins using MR images. In: Proceedings of the 28th IEEE EMBS annual international conference, New York, NY, 30 Aug–3 Sept 2006
26. Vistrails. Vistrails, University of Utah. Available: http://www.sci.utah.edu/cibc/software/
27. Barratt DC, Ariff BB, Humphries KN, Thom SAMG, Hughes AD (2004) Reconstruction and quantification of the carotid artery bifurcation from 3-D ultrasound images. IEEE Trans Med Imaging 23(5):567–583
28. Jialiang C, Shengzhang W, Wei Y, Guanghong D (2008) Computational fluid dynamics modeling of intracranial aneurysms. In: 2008 International conference on biomedical engineering and informatics
29. Swillens A, Segers P, Lovstakken L, Torp H. A comparison of multidimensional flow estimation techniques using computational fluid dynamics: speckle tracking versus vector Doppler. In: 2008 IEEE international ultrasonics symposium proceedings
30. Medis. Medis, Leiden, The Netherlands. Available at: http://www.medis.nl/
31. Musick M (2008) Mason researchers unveil new blood flow simulation system for diagnosing brain aneurysms. Mason Gazette, 9 June 2008
32. Ansys. Fluent, AnSYS. Available at: http://www.fluent.com/
33. Varghese SS, Frankel SH, Fischer PF (2007) Direct numerical simulation of stenotic flows. Part 1. Steady flow. J Fluid Mech 582:253–280
34. Varghese SS, Frankel SH, Fischer PF (2007) Direct numerical simulation of stenotic flows. Part 2. Pulsatile flow. J Fluid Mech 582:281–318
35. Deville MO, Fischer PF, Mund EH (2002) High-order methods for incompressible fluid flow. Cambridge University Press, Cambridge, UK
36. Choi G, Cheng CP, Wilson NM, Taylor CA (2009) Methods for quantifying three-dimensional deformation of arteries due to pulsatile and nonpulsatile forces: implications for the design of stents and stent grafts. Ann Biomed Eng 37(1):14–33
37. Humphrey JD, Taylor CA (2008) Intracranial and abdominal aortic aneurysms: similarities, differences, and need for a new class of computational models. Annu Rev Biomed Eng 10: 221–46
38. Tang BT, Cheng CP, Draney MT, Wilson NM, Tsao PS, Herfkens RJ, Taylor CA (2006) Abdominal aortic hemodynamics in young healthy adults at rest and during lower limb exercise: quantification using image-based computer modeling. Am J Physiol Heart Circ Physiol 291:H668–H676

39. Vignon-Clementel IE, Figueroa CA, Jansen KE, Taylor CA (2006) Outflow boundary conditions for three-dimensional finite element modeling of blood flow and pressure in arteries. Comput Methods Appl Mech Eng 195:29–32, 3776–3796
40. SYNERGY. Available: http://synergy2.sourceforge.net/
41. Garbey M, Hadri B, Hilford V, Karmonik C (2007) Parallel image-based hemodynamic simulator. Proceeding of ICSNC'07, workshop HPC-Bio07, France
42. Garbey M, Picard C (2007) Toward a general solution verification method for complex PDE problem with hands off coding. In: 45th AIAA Aerospace Sciences Meeting and Exhibit, 2007. AIAA-2007-1132
43. McDonald DA (1990) Blood flow in arteries, 3rd edn. Edward Arnold, London
44. Shyy W (2001) Moving boundaries in micro-scale biofluid dynamics. Appl Mech Rev 54(5):405–453
45. Ye T, Mittal R, Udaykumar HS, Shyy W (1999) An accurate Cartesian grid method for viscous incompressible flows with complex immersed boundaries. JCP 156:209–240
46. Arquis E, Caltagirone JP (1984) Sur les conditions hydrodynamiques au voisinage d'une interface milieu fluide-milieux poreux: application a la convection naturelle. CR Acad Sci Paris II 299:1–4
47. Schneider K, Farge M (2005) Numerical simuation of the transient flow behavior in tube bundles using a volume penalization method. J Fluids Struct 20:555–566
48. Angot P, Bruneau CH, Fabrie P (1999) A penalisation method to take into account obstacles in viscous flows. Numerische Mathematik 81:497–520
49. Chan TF, Vese LA (2001) Active contours without edges. IEE Transac Image Process 10(2):266–277
50. Osher S, Paragios N (2003) Geometric level set methods in imaging, vision and graphics, ISBN 0387954880. Springer, New York
51. Sethian JA (1999) Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. In: Ciarlet PG, Iserles A, Kohn RV, Wright MH (eds) Cambridge Monographs on Computational Mathematics (No. 3)
52. Barberou N, Garbey M, Hess M, Resch M, Rossi T, Toivanen J, Tromeur Dervout D (2003) Efficient metacomputing of elliptic linear and non-linear problems. J Parallel Distrib Comput 63:564–577
53. Garbey M, Tromeur Dervout D (2002) On some Aitken like acceleration of the Schwarz Method. Int J Numer Methods Fluids 40:1493–1513
54. Chorin AJ (1967) The numerical solution on the Navier–Stokes equations for an incompressible fluid. Bull Am Math Soc 73:928
55. David Keyes. Technologies and tools for high-performance distributed computing. In: Seventh workshop on the DOE, Advanced CompuTational Software (ACTS) collection, Aug 2006
56. Garbey M (2005) Acceleration of the Schwarz method for elliptic problems. SIAM J Sci Comput 26(6):1871–1893
57. Garbey M, Hadri B, Shyy W. Fast elliptic solver for incompressible Navier–Stokes flow and heat transfer problems on the grid. In: 43rd Aerospace sciences meeting and exhibit conference, Reno, Jan 2005, paper number: AIAA-2005-1386, p 12
58. Garbey M, Hadri B (2006) Toward real time image base CFD. In: 17th international conference on domain decomposition methods, July 2006, Austria
59. Garbey M, Pacull F (2007) A versatile incompressible Navier–Stokes solver for blood flow application. Int J Numer Methods Fluids 54(5):473–496
60. Hadri B (2008) A fast domain decomposition applied to blood flow simlulation. PhD Dissertation, Department of Computer Science, University of Houston, Mar 2008, 122 pp

Index