Nicholas Ayache
Hervé Delingette
Polina Golland
Kensaku Mori (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012

15th International Conference
Nice, France, October 2012
Proceedings, Part II

2 Part II

MICCAI

Springer

Lecture Notes in Computer Science 7511

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Nicholas Ayache   Hervé Delingette
Polina Golland   Kensaku Mori (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012

15th International Conference
Nice, France, October 1-5, 2012
Proceedings, Part II

Springer

Volume Editors

Nicholas Ayache
Hervé Delingette
Inria Sophia Antipolis
Project Team Asclepios
06902 Sophia Antipolis, France
E-mail: {nicholas.ayache, herve.delingette}@inria.fr

Polina Golland
MIT, CSAIL
Cambridge, MA 02139, USA
E-mail: polina@csail.mit.edu

Kensaku Mori
Nagoya University
Information and Communications Headquarters
Nagoya, 464-8603, Japan
E-mail: kensaku@is.nagoya-u.ac.jp

Preface

The 15th International Conference on Medical Image Computing and Computer Assisted Intervention, MICCAI 2012, was held in Nice, France, at the Acropolis Convention Center during October 1–5, 2012.

Over the past 14 years, the MICCAI conferences have become a premier international event with full articles of high standard, indexed by Pubmed, and annually attracting leading scientists, engineers and clinicians working at the intersection of sciences, technologies and medicine.

It is interesting to recall that the MICCAI conference series was formed in 1998 by the merger of CVRMed (Computer Vision, Virtual Reality and Robotics in Medicine), MRCAS (Medical Robotics and Computer Assisted Surgery) and VBC (Visualization in Biomedical Computing) conferences, and that the first CVRMed conference was held in Nice in April 1995. At that time the CVRMed conference was a single event and the proceedings, also published in Lecture Notes in Computer Science (LNCS), consisted of a single volume of 570 pages. In 2012 the MICCAI proceedings span three volumes and more than 2000 pages, and the conference was complemented by 32 MICCAI satellite events (workshops, challenges, tutorials) publishing their own proceedings, several of them in LNCS.

MICCAI contributions were selected through a rigorous reviewing process involving an international Program Committee (PC) of 100 specialists coordinated by a Program Chair and 2 Program Co-chairs from 3 continents. Decisions were based on anonymous reviews made by 913 expert reviewers. The process was double blind as authors did not know the names of the PC members/reviewers evaluating their papers, and the PC members/reviewers did not know the names of the authors of the papers they were evaluating.

We received 781 submissions and after the collection of over 3000 anonymous review forms, the final selection was prepared during a 2-day meeting in Nice (12–13 May 2012) attended by 50 PC members. They finalized the acceptation of 252 papers (i.e., acceptance rate of 32%) and also prepared a short list of candidate papers for plenary presentations. The accepted contributions came from 21 countries and 5 continents: about 50% from North America (40% USA and 8% Canada), 40% from Europe (mainly from France, Germany, the UK, Switzerland and The Netherlands), and 10% from Asia and the rest of the world.

All accepted papers were presented during 6 poster sessions of 90 minutes with the option, this year for the first time, of displaying additional dynamic material on large screens during the whole poster session. In addition, a subset of 37 carefully selected papers (mainly chosen among the short list of candidate papers recommended by PC members) were presented during 7 single-track plenary oral sessions.

Prof. Alain Carpentier, President of the French Academy of Sciences, was the Honored Guest of MICCAI 2012 for his pioneering and visionary role in several of the domains covered by MICCAI. Prof. Carpentier addressed the audience during the opening ceremony along with Prof. Michel Cosnard, the CEO of Inria, and introduced one the keynote lectures.

Prof. Jacques Marescaux, director of the Strasbourg IHU (Institut Hospitalo-Universitaire) delivered the keynote lecture "Surgery for Life Innovation: Information Age and Robotics" and Prof. Michel Haïssaguerre, director of the Bordeaux IHU, delivered the keynote lecture "Preventing Sudden Cardiac Death: Role of Structural and Functional Imaging". Both of these lectures were outstanding and inspiring.

The conference would not have been possible without the commitment and hard work of many people whom we want to thank wholeheartedly:

- The 100 Program Committee members and 913 scientific reviewers, listed in this book, who worked closely with us and prepared many written reviews and recommendations for acceptance or rejection,
- Xavier Pennec as the Chair for the organization of the 32 satellite events (workshops, challenges, tutorials) with the assistance of Tobias Heimann, Kilian Pohl and Akinobu Shimizu as Co-chairs, and all the organizers of these events,
- Agnès Cortell as the Local Organization Chair, who successfully coordinated all the details of the organization of the event with the support of a local organizing team (composed of Marc Barret, Grégoire Malandain, Xavier Pennec, Maxime Sermesant and two of us), several Inria services (involving heavily Odile Carron and Matthieu Oricelli), and the MCI company,
- Maxime Sermesant as MICCAI Website Chair,
- Grégoire Malandain for the new organization of posters including digital screens,
- Isabelle Strobant for the organization of the PC meeting in Nice, the invitations of the MICCAI guests, and her constant support during the preparation of the event,
- Gérard Giraudon, director of Inria in Sophia Antipolis, for his constant support,
- Sebastien Ourselin for his help in coordinating industrial sponsorship,
- All students and engineers (mainly from Asclepios and Athena Inria teams) who helped with the scientific and local organization,
- Emmanuelle Viau, who coordinated the team at MCI including in particular Thibault Claisse and Thibault Lestiboudois,
- Jim Duncan as the President of the MICCAI Society and its board of directors who elected MICCAI 2012 to be held in Nice,
- Janette Wallace, Johanne Guillemette and Chris Wedlake for the liaison with the MICCAI Society,
- James Stewart for his precious help with the Precision Conference System,
- All our industrial and institutional sponsors and partners for their fantastic support of the conference.

Finally, we would like to thank all the MICCAI 2012 attendees who came to Nice from 34 countries from all around the world, and we look forward to meeting them again at MICCAI 2013 in Nagoya, Japan, at MICCAI 2014 in Cambridge, Massachusetts, USA and at MICCAI 2015 in Munich, Germany.

October 2012

Nicholas Ayache
Hervé Delingette
Polina Golland
Kensaku Mori

Accepted MICCAI 2012 Papers

by Clinical Theme

- Others 9%
- Oncology 6%
- Musculo-Skeleton 8%
- NeuroImaging 35%
- Abdominal 22%
- Biological 8%
- CardioVascular 16%

by Technical Theme

- Computed-Aided Diagnosis 15%
- Planning and Simulation 8%
- Segmentation 31%
- Statistical Analysis 13%
- Registration 13%
- Acquisition & Enhancement 12%
- Robotics & Tracking 7%
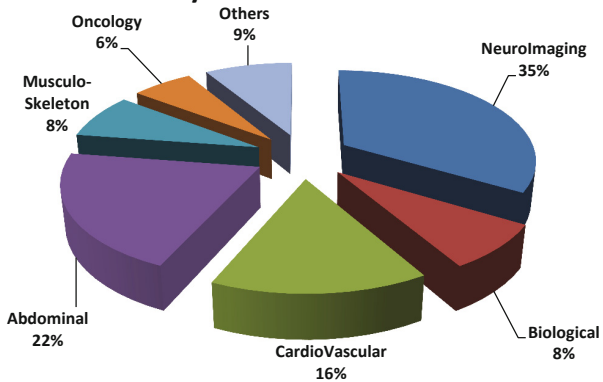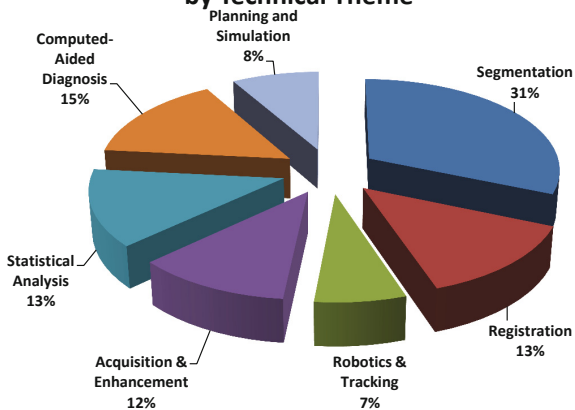
by Country of First Author

- Others in Europe 6%
- Asia 6%
- Rest of the world 4%
- Switzerland 4%
- United Kingdom 10%
- United States 40%
- Germany 11%
- France 12%
- Canada 8%

Organization

General Chair

Nicholas Ayache   Inria, Sophia Antipolis, France

Program Chair and Co-chairs

Hervé Delingette	Inria, Sophia Antipolis, France
Polina Golland	MIT, Cambridge, USA
Kensaku Mori	Nagoya University, Nagoya, Japan

Workshops, Tutorials and Challenges Chair and Co-chairs

Xavier Pennec	Inria, Sophia Antipolis, France
Tobias Heimann	Cancer Research Center, Heidelberg, Germany
Kilian Pohl	University of Pennsylvania, Philadelphia, USA
Akinobu Shimizu	Tokyo University of A&T, Tokyo, Japan

MICCAI Society, Board of Directors

James Duncan (President)	Yale University, USA
Gabor Fichtinger (Treasurer)	Queen's University, Canada
Alison Noble (Exec. Director)	University of Oxford, UK
Sebastien Ourselin (Secretary)	University College London, UK
Nicholas Ayache	Inria Sophia Antipolis, France
Polina Golland	MIT, USA
David Hawkes	University College London, UK
Kensaku Mori	Nagoya University, Japan
Wiro Niessen	Erasmus MC, The Netherlands
Xavier Pennec	Inria Sophia Antipolis, France
Daniel Rueckert	Imperial College London, UK
Dinggang Shen	University North Carolina, USA
William Wells	Harvard Medical School, USA

Consultants to Board

Alan Colchester	University of Kent, UK
Terry Peters	Robarts Research Institute, Canada
Richard Robb	Mayo Clinic College of Medicine, USA

Program Committee

Purang Abolmaesumi	University of British Columbia, Canada
Daniel Alexander	University College London, UK
Amir Amini	University Louisville, USA
Elsa Angelini	Télécom ParisTech, France
Stephen Aylward	Kitware, USA
Christian Barillot	CNRS, France
Wolfgang Birkfellner	Medical University of Vienna, Austria
Oscar Camara	University Pompeu Fabra, Spain
Albert Chung	HKUST, Hong Kong
Ela Claridge	University of Birmingham, UK
Patrick Clarysse	University of Lyon, France
Louis Collins	McGill University, Canada
Olivier Colliot	ICM-CNRS, France
Dorin Comaniciu	Siemens, USA
Stéphane Cotin	Inria, France
Antonio Criminisi	Microsoft Research, UK
Christos Davatzikos	University of Pennsylvania, USA
Marleen de Bruijne	Erasmus MC, The Netherlands
Rachid Deriche	Inria, France
James Duncan	University of Yale, USA
Philip Edwards	Imperial College London, UK
Gabor Fichtinger	Queen's University, Canada
Bernd Fischer	University of Luebeck, Germany
Thomas Fletcher	University of Utah, USA
Alejandro Frangi	University Pompeu Fabra, Spain
Jim Gee	University of Pennsylvania, USA
Guido Gerig	University of Utah, USA
Leo Grady	Siemens, USA
Hayit Greenspan	Tel Aviv University, Israel
Gregory Hager	John's Hopkins University, USA
Heinz Handels	University of Luebeck, Germany
Matthias Harders	ETH Zurich, Switzerland
Nobuhiko Hata	Harvard Medical School, USA
David Hawkes	University College London, UK
Tobias Heimann	DKFZ, Germany
Ameet Jain	Philips, USA
Pierre Jannin	INSERM, France
Marie-Pierre Jolly	Siemens, USA
Leo Joskowicz	University of Jerusalem, Israel
Ioannis Kakadiaris	University of Houston, USA
Nico Karssemeijer	Radboud University, The Netherlands
Ron Kikinis	Harvard Medical School, USA

Regis Vaillant	General Electric, France
Bram van Ginneken	Radboud University, The Netherlands
Koen Van Leemput	Harvard Medical School, USA
Baba Vemuri	University of Florida, USA
Ragini Verma	University of Pennsylvania, USA
Simon Warfield	Harvard Medical School, USA
Jurgen Weese	Philips, Germany
Wolfgang Wein	Technical University of Munich, Germany
William Wells	Harvard Medical School, USA
Carl-Fredrik Westin	Harvard Medical School, USA
Guang Zhong Yang	Imperial College London, UK
Laurent Younes	John's Hopkins University, USA
Alistair Young	University of Auckland, New Zealand

Organizing Institution

This event was organized by Inria, the French Research Institute for Computer Science and Applied Mathematics.

Local Organizing Committee

Agnès Cortell	Inria, Sophia Antipolis, France
Nicholas Ayache	Inria, Sophia Antipolis, France
Marc Barret	Inria, Sophia Antipolis, France
Hervé Delingette	Inria, Sophia Antipolis, France
Grégoire Malandain	Inria, Sophia Antipolis, France
Xavier Pennec	Inria, Sophia Antipolis, France
Maxime Sermesant	Inria, Sophia Antipolis, France
Isabelle Strobant	Inria, Sophia Antipolis, France

Liaison with the MICCAI Society

Janette Wallace	Robarts Research Institute, London, Canada
Johanne Guillemette	Robarts Research Institute, London, Canada

Official Partners

Institut Océanographique de Monaco
Région Provence Alpes Côte d'Azur
Ville de Nice

Sponsors

Gold Sponsors	GE HealthCare
	Philips
	Siemens
	Canon Median
Silver Sponsors	ERC MedYMA
	Medtronic
Bronze Sponsors	Aviesan
	Dosisoft
	IHU Strasbourg
	IRCAD France
	Kitware
	Microsoft Research

Exhibitors

Camelot Biomedial systems	Claron Technology
Elsevier	NDI
Springer	Ultrasonix
VSG Visualization Sciences Group	

Reviewers

Abramoff, Michael	Andres, Bjoern
Acar, Burak	Antani, Sameer
Achterberg, Hakim	Anwander, Alfred
Acosta-Tamayo, Oscar	Arbel, Tal
Adluru, Nagesh	Arimura, Hidetaka
Aganj, Iman	Arridge, Simon R.
Ahmadi, Seyed-Ahmad	Ashburner, John
Aja-Fernández, Santiago	Astley, Sue
Akcakaya, Mehmet	Atkinson, David
Akhondi-Asl, Alireza	Audette, Michel
Alander, Jarmo	Augustine, Kurt
Alberola-López, Carlos	Auvray, Vincent
Alexander, Andrew	Avants, Brian
Ali, Sahirzeeshan	Avila, Rick
Aljabar, Paul	Awate, Suyash
Allain, Baptiste	Axel, Leon
Allassonnière, Stephanie	Ayad, Maria
Amini, Amir	Bach Cuadra, Meritxell
An, Jungha	Baddeley, David
Anderson, Adam	Baghani, Ali
Andersson, Jesper	Baka, Nora

Balicki, Marcin
Ballerini, Lucia
Baloch, Sajjad
Barbu, Adrian
Barmpoutis, Angelos
Barratt, Dean
Barré, Arnaud
Basavanhally, Ajay
Batmanghelich, Nematollah
Bazin, Pierre-Louis
Beichel, Reinhard
Belongie, Serge
Ben Ayed, Ismail
Benajiba, Yassine
Benali, Habib
Bengtsson, Ewert
Bergeles, Christos
Berger, Marie-Odile
Bergtholdt, Martin
Berks, Michael
Bernal, Jorge Luis
Bernard, Olivier
Bernus, Olivier
Betrouni, Nacim
Bezy-Wendling, Johanne
Bhatia, Kanwal
Bhotika, Rahul
Biesdorf, Andreas
Bilgazyev, Emil
Bilgic, Berkin
Bishop, Martin
Bismuth, Vincent
Blaschko, Matthew
Bloch, Isabelle
Bloy, Luke
Blum, Tobias
Bogunovic, Hrvoje
Boisvert, Jonathan
Bosch, Johan
Bossa, Matias Nicolas
Bouarfa, Loubna
Bouix, Sylvain
Boukerroui, Djamal
Bourgeat, Pierrick
Bovendeerd, Peter

Brady, Michael
Breitenreicher, Dirk
Brock, Kristy
Brost, Alexander
Brun, Caroline
Burlina, Philippe
Butakoff, Constantine
Buvat, Irène
Caan, Matthan
Cahill, Nathan
Cai, Weidong
Cameron, Bruce
Camp, Jon
Cardenas, Valerie
Cardenes, Ruben
Cardoso, Manuel Jorge
Carmichael, Owen
Carson, Paul
Castaeda, Victor
Castro-Gonzalez, Carlos
Cathier, Pascal
Cattin, Philippe C.
Celebi, M. Emre
Cetingul, Hasan Ertan
Chakravarty, M. Mallar
Chan, Raymond
Chappelow, Jonathan
Chaux, Caroline
Chen, Elvis C. S.
Chen, Terrence
Chen, Ting
Chen, Xinjian
Chen, Yen-Wei
Chen, Yunmei
Cheng, Guang
Cheng, Jian
Cheriet, Farida
Chintalapani, Gouthami
Chinzei, Kiyoyuki
Chitphakdithai, Nicha
Chou, Yiyu
Chowdhury, Ananda
Christensen, Gary
Chu, Chia-Yueh Carlton
Chung, Moo K.

Chupin, Marie
Cinquin, Philippe
Ciofolo, Cybele
Ciompi, Francesco
Ciuciu, Philippe
Clark, Alys
Clarkson, Matthew
Cleary, Kevin
Clerc, Maureen
Clouchoux, Cédric
Cloutier, Guy
Combès, Benoît
Commowick, Olivier
Cootes, Tim
Corso, Jason
Coudiere, Yves
Coulon, Olivier
Coupe, Pierrick
Cowan, Brett
Crimi, Alessandro
Crum, William
Cui, Xinyi
Cuingnet, Remi
D'Alessandro, Brian
Daga, Pankaj
Dahl, Anders L.
Dai, Yakang
Daoud, Mohammad
Darkner, Sune
Darvann, Tron
Darzi, Ara
Dauguet, Julien
Dawant, Benoit
De Craene, Mathieu
Debbaut, Charlotte
Dehghan, Ehsan
Deligianni, Fani
Delong, Andrew
Demiralp, Cagatay
Demirci, Stefanie
Deng, Xiang
Dennis, Emily
Dequidt, Jeremie
Desbat, Laurent
Descoteaux, Maxime

Desvignes, Michel
Dewan, Maneesh
D'Haese, Pierre-François
DiBella, Edward
Diciotti, Stefano
Dijkstra, Jouke
Dikici, Engin
DiMaio, Simon
Ding, Kai
Dinten, Jean-Marc
Doessel, Olaf
Doignon, Christophe
Dojat, Michel
Dong, Bin
Donner, René
Douglas, Tania
Douiri, Abdel
Dowling, Jason
Doyle, Scott
Drangova, Maria
Drechsler, Klaus
Drobnjak, Ivana
Duan, Qi
Duchateau, Nicolas
Duchesnay, Edouard
Duchesne, Simon
Duriez, Christian
Durrleman, Stanley
Dzyubachyk, Oleh
Eagleson, Roy
Ebbers, Tino
Ecabert, Olivier
Ehrhardt, Jan
Elad, Michael
El-Baz, Ayman
Elen, An
Eleonora, Fornari
Elhawary, Haytham
El-Zehiry, Noha
Ennis, Daniel
Enquobahrie, Andinet
Erdt, Marius
Eskandari, Hani
Eskildsen, Simon
Eslami, Abouzar

Heberlein, Keith
Heckemann, Rolf
Heinrich, Mattias Paul
Hellier, Pierre
Heng, Pheng Ann
Hennemuth, Anja
Herlambang, Nicholas
Hernandez, Monica
Hipwell, John
Hirano, Yasushi
Hoffmann, Kenneth
Holmes, David
Hontani, Hidekata
Hoogendoorn, Corné
Hornegger, Joachim
Howe, Robert
Hsu, Li-Yueh
Hu, Yipeng
Hu, Zhihong
Huang, Heng
Huang, Junzhou
Huang, Rui
Huang, Wei
Huang, Xiaolei
Hudelot, Céline
Huisman, Henkjan
Humbert, Ludovic
Hurdal, Monica
Hyde, Damon
Iakovidis, Dimitris
Iglesias, Juan Eugenio
Imiya, Atsushi
Ingalhalikar, Madhura
Ionasec, Razvan
Irfanoglu, Mustafa Okan
Isgum, Ivana
Ishikawa, Hiroshi
Jacob, Mathews
Jacobs, Colin
Jahanshad, Neda
Janoos, Firdaus
Janowczyk, Andrew
Jbabdi, Saad
Jenkinson, Mark
Jerebko, Anna

Jian, Bing
Jiang, Tianzi
Jiang, Yifeng
Jomier, Julien
Jordan, Petr
Joshi, Anand
Joshi, Sarang
Jurrus, Elizabeth
Kabus, Sven
Kachelrie, Marc
Kadoury, Samuel
Kainmueller, Dagmar
Kallenberg, Michiel
Kamen, Ali
Kanade, Takeo
Kapoor, Ankur
Kapur, Tina
Karamalis, Athanasios
Karemore, Gopal
Krsnäs, Andreas
Karwoski, Ron
Kaster, Frederik
Katouzian, Amin
Kawata, Yoshiki
Kaynig, Verena
Kazanzides, Peter
Keeve, Erwin
Kelm, Michael
Kerrien, Erwan
Kezele, Irina
Khan, Ali R.
Kherif, Ferath
Khurd, Parmeshwar
Kim, Boklye
Kim, Kio
Kim, Minjeong
Kindlmann, Gordon
King, Andrew
Kiraly, Atilla
Kirchberg, Klaus
Kitasaka, Takayuki
Klein, Arno
Klein, Jan
Klein, Martina
Klein, Stefan

Klein, Tassilo
Klinder, Tobias
Klöppel, Stefan
Knoesche, Thomas R.
Knoll, Alois
Kobayahsi, Etsuko
Kohannim, Omid
Kohlberger, Timo
Kohli, Pushmeet
Konukoglu, Ender
Kozerke, Sebastian
Krissian, Karl
Kroenke, Christopher
Kruggel, Frithjof
Kumar, Rajesh
Kumar, Ritwik
Kurkure, Uday
Kuroda, Yoshihiro
Kwok, Ka-Wai
Kwon, Dongjin
Kybic, Jan
Ladikos, Alexander
Laine, Andrew
Lalande, Alain
Lalys, Florent
Lamecker, Hans
Landman, Bennett
Lango, Thomas
Langs, Georg
Lapeer, Rudy
Laporte, Catherine
Lartizien, Carole
Lasso, Andras
Lauze, François
Law, Max W.K.
Le Montagner, Yoan
Le, Yen
Lee, Angela
Lee, John
Lee, Junghoon
Lee, Su-Lin
Lee, Tim
Lekadir, Karim
Lelieveldt, Boudewijn
Lensu, Lasse

Leow, Alex
Lepore, Natasha
Lesage, David
Leung, Kelvin
Li, Bo
Li, Chunming
Li, Fuhai
Li, Gang
Li, Hongsheng
Li, Kaiming
Li, Ming
Li, Yang
Liao, Hongen
Liao, Rui
Liao, Shu
Liebling, Michael
Lindseth, Frank
Ling, Haibin
Linguraru, Marius George
Linte, Cristian
Litjens, Geert
Liu, Huafeng
Liu, Jiamin
Liu, Manhua
Liu, Meizhu
Liu, Sheena
Liu, Tianming
Liu, Xiaofeng
Liu, Xiaoxiao
Liu, Zhao
Lo, Pechin
Loeckx, Dirk
Loew, Murray
Lohmann, Gabriele
Lombaert, Herve
Loog, Marco
Lötjönen, Jyrki
Lu, Chao
Lu, Le
Lu, Xiaoguang
Luboz, Vincent
Lucas, Blake
Lui, Lok Ming
Luo, Yishan
Lynch, John

Ma, YingLiang
Machiraju, Raghu
MacLeod, Robert
Madany Mamlouk, Amir
Maddah, Mahnaz
Magee, Derek
Magnotta, Vincent
Maier-Hein, Lena
Malandain, Grégoire
Manduca, Armando
Mani, Meena
Manjón, José V.
Manniesing, Rashindra
Mansi, Tommaso
Manzke, Robert
Marchal, Maud
Marsland, Stephen
Martí, Robert
Masamune, Ken
Mattes, Julian
Maurel, Pierre
Mavroforakis, Michael
McClelland, Jamie
McCormick, Matthew
Medrano-Gracia, Pau
Meine, Hans
Meinzer, Hans-Peter
Meisner, Eric
Mekada, Yoshito
Melbourne, Andrew
Mertins, Alfred
Metz, Coert
Meyer, Chuck
Meyer, François
Michailovich, Oleg
Michel, Fabrice
Mihalef, Viorel
Miller, James
Modat, Marc
Modersitzki, Jan
Mohamed, Ashraf
Monaco, James
Montillo, Albert
Moore, John
Moradi, Mehdi

Mory, Benoit
Müller, Henning
Murgasova, Maria
Murphy, Keelin
Mylonas, George
Najman, Laurent
Nakajima, Yoshikazu
Nakamura, Ryoichi
Nassiri-Avanaki, Mohammad-Reza
Negahdar, Mohammadjavad
Negahdar, Mohammadreza
Nekolla, Stephan
Neumuth, Thomas
Ng, Bernard
Nichols, Thomas
Nicolau, Stéphane
Nie, Jingxin
Niederer, Steven
Niethammer, Marc
Noble, Jack
Noël, Peter
Nolte, Lutz
Nordsletten, David
Nuyts, Johan
O'Brien, Kieran
Oda, Masahiro
O'Donnell, Lauren
O'Donnell, Thomas
Oguz, Ipek
Okada, Kazunori
Olabarriaga, Silvia
Olesch, Janine
Oliver, Arnau
Olmos, Salvador
Oost, Elco
Orihuela-Espina, Felipe
Orkisz, Maciej
Otake, Yoshito
Ou, Yangming
Pace, Danielle
Padfield, Dirk
Padoy, Nicolas
Palaniappan, Kannappan
Pallavaram, Srivatsan
Panagiotaki, Eleftheria

Paniagua, Beatriz
Paolillo, Alfredo
Papademetris, Xenios
Papadopoulo, Theo
Park, Mi-Ae
Parthasarathy, Vijay
Passat, Nicolas
Pasternak, Ofer
Patriciu, Alexandru
Paul, Perrine
Paulsen, Keith
Paulsen, Rasmus
Pauly, Olivier
Pavlidis, Ioannis
Pearlman, Paul
Pedemonte, Stefano
Peitgen, Heinz-Otto
Pekar, Vladimir
Peng, Hanchuan
Penney, Graeme
Pernus, Franjo
Perperidis, Antonios
Perrot, Matthieu
Peters, Amanda
Petersen, Jens
Petitjean, Caroline
Peyrat, Jean-Marc
Peyré, Gabriel
Pham, Dzung
Phlypo, Ronald
Piella, Gemma
Pitiot, Alain
Pizaine, Guillaume
Pizer, Stephen
Platel, Bram
Podder, Tarun
Poignet, Philippe
Poline, Jean-Baptiste
Polzehl, Joerg
Pontre, Beau
Poot, Dirk
Popovic, Aleksandra
Poupon, Cyril
Poynton, Clare
Pozo, José Maria

Prasad, Gautam
Prastawa, Marcel
Pratt, Philip
Prima, Sylvain
Prince, Jerry
Punithakumar, Kumaradevan
Puy, Gilles
Qazi, Arish A.
Qian, Zhen
Quellec, Gwenole
Radau, Perry
Radeva, Petia
Radulescu, Emil
Rahman, Md Mahmudur
Raj, Ashish
Rajagopalan, Srinivasan
Rajagopalan, Vidya
Rajpoot, Nasir
Rangarajan, Anand
Rasoulian, Abtin
Rathi, Yogesh
Ratnanather, Tilak
Ravishankar, Saiprasad
Reichl, Tobias
Reilhac-Laborde, Anthonin
Rettmann, Maryam
Reuter, Martin
Reyes, Mauricio
Reyes-Aldasoro, Constantino
Rhode, Kawal
Ribbens, Annemie
Richa, Rogerio
Riddell, Cyrill
Ridgway, Gerard
Riklin Raviv, Tammy
Risholm, Petter
Risser, Laurent
Rit, Simon
Rittscher, Jens
Rivaz, Hassan
Riviere, Cameron
Riviere, Denis
Roche, Alexis
Rohkohl, Christopher
Rohling, Robert

Rohr, Karl
Rousseau, François
Roysam, Badrinath
Ruehaak, Jan
Russakoff, Daniel
Rusu, Mirabela
Ruthotto, Lars
Sabczynski, Jörg
Sadeghi-Naini, Ali
Sadowsky, Ofri
Saha, Punam Kumar
Salvado, Olivier
San Jose Estepar, Raul
Sanchez, Clarisa
Sanderson, Allen
Sands, Greg
Sarrut, David
Sarry, Laurent
Savadjiev, Peter
Scherer, Reinhold
Scherrer, Benoit
Schindelin, Johannes
Schmidt, Michael
Schmidt-Richberg, Alexander
Schneider, Caitlin
Schneider, Torben
Schoonenberg, Gert
Schultz, Thomas
Schweikard, Achim
Sebastian, Rafael
Seiler, Christof
Serre, Thomas
Seshamani, Sharmishtaa
Shah, Shishir
Shamir, Reuben R.
Shen, Li
Shen, Tian
Shi, Feng
Shi, Kuangyu
Shi, Pengcheng
Shi, Yonggang
Shi, Yonghong
Shi, Yubing
Sijbers, Jan
Simaan, Nabil

Simonyan, Karen
Simpson, Amber
Simpson, Ivor
Singh, Maneesh
Singh, Nikhil
Singh, Vikas
Sinkus, Ralph
Siqueira, Marcelo
Sjöstrand, Karl
Slabaugh, Greg
Slagmolen, Pieter
Smal, Ihor
Smeets, Dirk
Soeller, Christian
Sofka, Michal
Soler, Luc
Song, Sang-Eun
Song, Xubo
Sonka, Milan
Srensen, Lauge
Sotiras, Aristeidis
Sparks, Rachel
Sporring, Jon
Staal, Joes
Staring, Marius
Staroswiecki, Ernesto
Stehle, Thomas
Stewart, James
Stolka, Philipp
Stoyanov, Danail
Styles, Iain
Subramanian, Navneeth
Suinesiaputra, Avan
Sundar, Hari
Suthau, Tim
Suzuki, Kenji
Syeda-Mahmood, Tanveer
Szczerba, Dominik
Tagare, Hemant
Tahmasebi, Amir
Tai, Xue-Cheng
Tannenbaum, Allen
Tanner, Christine
Tao, Xiaodong
Tasdizen, Tolga

Tavakoli, Vahid
Taylor, Zeike
Thévenaz, Philippe
Thiriet, Marc
Tiwari, Pallavi
Tobon-Gomez, Catalina
Toews, Matthew
Tohka, Jussi
Tokuda, Junichi
Tosun, Duygu
Toth, Robert
Toussaint, Nicolas
Tristán-Vega, Antonio
Tsekos, Nikolaos V.
Turaga, Srinivas
Tustison, Nicholas
Uchiyama, Yoshikazu
Udupa, Jayaram K.
Unal, Gozde
Uzunbas, Mustafa
van Assen, Hans
van der Geest, Rob
van der Lijn, Fedde
van Rikxoort, Eva
van Stralen, Marijn
van Walsum, Theo
Vannier, Michael
Varoquaux, Gael
Vegas-Sánchez-Ferrero, Gonzalo
Venkataraman, Archana
Vercauteren, Tom
Vialard, François-Xavier
Vignon, François
Villain, Nicolas
Villard, Pierre-Frédéric
Vincent, Nicole
Visentini-Scarzanella, Marco
Visvikis, Dimitris
Viswanath, Satish
Vitanovski, Dime
Vogel, Jakob
Voigt, Ingmar
von Berg, Jens
Voros, Sandrine
Vos, Pieter

Vosburgh, Kirby
Vrooman, Henri
Vrtovec, Tomaz
Wachinger, Christian
Waechter-Stehle, Irina
Wahle, Andreas
Waldman, Lew
Wang, Chaohui
Wang, Fei
Wang, Hongzhi
Wang, Hui
Wang, Lejing
Wang, Li
Wang, Liansheng
Wang, Peng
Wang, Qian
Wang, Song
Wang, Vicky
Wang, Yalin
Wang, Yang
Wang, Ying
Wanyu, Liu
Warfield, Simon
Wassermann, Demian
Weber, Stefan
Wee, Chong-Yaw
Wei, Liu
Weiskopf, Nikolaus
Wells, William
Wels, Michael
Werner, Rene
Whitaker, Ross
Whitmarsh, Tristan
Wiles, Andrew
Wirtz, Stefan
Wittek, Adam
Wolf, Ivo
Wolz, Robin
Wörz, Stefan
Wu, Guorong
Wu, Wen
Wu, Xiaodong
Xenos, Michalis
Xie, Jun
Xiong, Guanglei

Xu, Jun
Xu, Lei
Xu, Sheng
Xu, Xiayu
Xue, Hui
Xue, Zhong
Yan, Pingkun
Yan, Zhennan
Yang, Fei
Yang, Lin
Yang, Xiaofeng
Yang, Xiaoyun
Yaniv, Ziv
Yao, Jianhua
Yap, Pew-Thian
Yaqub, Mohammad
Ye, Dong Hye
Yener, Bülent
Yeniaras, Erol
Yeo, B.T. Thomas
Yin, Zhaozheng
Ying, Leslie
Yoo, Terry
Yoshida, Hiro
Yotter, Rachel
Yushkevich, Paul
Zagorchev, Lyubomir
Zahiri Azar, Reza
Zaidi, Habib
Zeng, Wei

Zhan, Liang
Zhan, Yiqiang
Zhang, Chong
Zhang, Daoqiang
Zhang, Honghai
Zhang, Hui
Zhang, Jingdan
Zhang, Pei
Zhang, Shaoting
Zhao, Fei
Zheng, Guoyan
Zheng, Yefeng
Zheng, Yuanjie
Zhong, Hua
Zhong, Lin
Zhou, Jinghao
Zhou, Luping
Zhou, S. Kevin
Zhou, X. Sean
Zhou, Xiaobo
Zhou, Yan
Zhu, Hongtu
Zhu, Ning
Zhu, Yuemin
Zhuang, Xiahai
Zijdenbos, Alex
Zikic, Darko
Zion, Tse
Zollei, Lilla
Zwiggelaar, Reyer

Awards Presented at MICCAI 2011, Toronto

MICCAI Society Enduring Impact Award Sponsored by Philips: The Enduring Impact Award is the highest award of the MICCAI Society. It is a career award for continued excellence in the MICCAI research field. The 2011 Enduring Impact Award was presented to *Chris Taylor*, Manchester University, UK.

MICCAI Society Fellowships: MICCAI Fellowships are bestowed annually on a small number of senior members of the society in recognition of substantial scientific contributions to the MICCAI research field and service to the MICCAI community. In 2011, fellowships were awarded to:

- *Christian Barillot* (IRISA-CNRS, France)
- *Gabor Fichtinger* (Queens University, Canada)
- *Jerry Prince* (Johns Hopkins University, USA)

Medical Image Analysis Journal Award Sponsored by Elsevier: *Ola Friman*, for the article entitled: "Probabilistic 4D Blood Flow Tracking and Uncertainty Estimation", co-authored by: Ola Friman, Anja Hennemuth, Andreas Harloff, Jelena Bock, Michael Markl, and Heinz-Otto Peitgen

Best Paper in Computer-Assisted Intervention Systems and Medical Robotics, Sponsored by Intuitive Surgical Inc.: *Jay Mung*, for the article entitled "A Non-disruptive Technology for Robust 3D Tool Tracking for Ultrasound-Guided Interventions", co-authored by: Jay Mung, Francois Vignon, and Ameet Jain.

MICCAI Young Scientist Awards: The Young Scientist Awards are stimulation prizes awarded for the best first authors of MICCAI contributions in distinct subject areas. The nominees had to be full-time students at a recognized university at, or within, two years prior to submission. The 2011 MICCAI Young Scientist Awards were given to:

- *Mattias Heinrich* for his paper entitled "Non-local Shape Descriptor: A New Similarity Metric for Deformable Multi-modal Registration"
- *Tommaso Mansi* for his paper entitled "Towards Patient-Specific Finite-Element Simulation of Mitral Clip Procedure"
- *Siyang Zuo* for his paper entitled "Nonmetalic Rigid-Flexible Outer Sheath with Pneumatic Shapelocking Mechanism and Double Curvature Structure"
- *Christof Seiler* for his paper entitled "Geometry-Aware Multiscale Image Registration via OBB Tree-Based Polyaffine Log-Demons"
- *Ting Chen* for her paper entitled "Mixture of Segmenters with Discriminative Spatial Regularization and Sparse Weight Selection"

Table of Contents – Part II

Cardiovascular Imaging: Planning, Intervention and Simulation

Image Registration I

NeuroImage Analysis I

Diffusion Weighted Imaging

Image Segmentation II

Cardiovascular Imaging II

Computer-Assisted Interventions and Robotics II

Image Registration: New Methods and Results

Automatic Multi-model-Based Segmentation of the Left Atrium in Cardiac MRI Scans

Dominik Kutra[1], Axel Saalbach[1], Helko Lehmann[1], Alexandra Groth[1],
Sebastian P.M. Dries[1], Martin W. Krueger[2], Olaf Dössel[2], and Jürgen Weese[1]

[1] Philips Research Laboratories, Hamburg, Germany
[2] Institute of Biomedical Engineering, Karlsruhe Institute of Technology (KIT),
Karlsruhe, Germany

Abstract. Model-based segmentation approaches have been proven to
produce very accurate segmentation results while simultaneously provid-
ing an anatomic labeling for the segmented structures. However, variations
of the anatomy, as they are often encountered e.g. on the drainage pattern
of the pulmonary veins to the left atrium, cannot be represented by a single
model. Automatic model selection extends the model-based segmentation
approach to handling significant variational anatomies without user inter-
action. Using models for the three most common anatomical variations of
the left atrium, we propose a method that uses an estimation of the local
fit of different models to select the best fitting model automatically. Our
approach employs the support vector machine for the automatic model se-
lection. The method was evaluated on 42 very accurate segmentations of
MRI scans using three different models. The correct model was chosen in
88.1 % of the cases. In a second experiment, reflecting average segmen-
tation results, the model corresponding to the clinical classification was
automatically found in 78.0 % of the cases.

1 Introduction

With a prevalence of 0.4% to 1%, atrial fibrillation (AF) is the most common
cardiac arrhythmia in the USA and in Europe [1] and can lead to severe, life-
threatening conditions like stroke. Catheter ablation procedures aim at the elec-
trical isolation of reentry pathways and ectopic foci that are causing AF in the
atrial tissue. Triggering foci are often found in the pulmonary veins (PVs) of
the left atrium (LA) [2]. The knowledge of the anatomy of the LA is crucial
for successful procedures as it enables accurate planning of ablation lines and
guidance during the procedure.

With the advent of ablation procedures to treat AF by the isolation of the
PVs, interest in the assessment of the PV configuration rose. *Typically*, the LA
is joined by two PVs on each side through individual ostia: left superior and left
inferior PV (LSPV, LIPV) on the left side of the LA; right superior and right
inferior PV (RSPV, RIPV) on the right. This, *typical* pattern is found in about
60 % of the patients [3]. The most frequent variation on the left side of the LA
is the common left trunk (CLT). Both left PVs merge into a single trunk in the

proximity of the LA body. With a distance from the ostium to the bifurcation shorter than 1 cm to 1.5 cm, depending on the study [3,4], it is called short CLT and occurs in 10 % of the patients while the CLT is found in 4 % to 8 % of the patients. Kaseno et al. suggest a typical configuration as well as the short CLT to be normal [4]. On the right side, the right middle PV pattern (RMPV), which is characterized by an accessory PV joining the LA body with a separate ostium, is found in 13 % to 24 % of the patients. The presence of concurrent variations on both sides of the LA is rarely reported.

Using model-based segmentation, very accurate segmentation results have been demonstrated in the past for the heart [5,6,7] in general and the LA in particular [8]. However, using only a single model, variations in the anatomy of the LA cannot be reflected and thus, the segmentation is inaccurate. Different approaches emerged from the necessity to account for the variational anatomy while preserving the advantages of model-based segmentation. Zheng et al. presented a part based model approach, adapting the chamber with the LAA and the four major PVs as individual parts on C-arm CT datasets [9]. In the adapted state, the parts are joined, enabling the approach to represent both, two individual PVs on the left as well as the CLT pattern. Hanna et al. introduced a hybrid method combining model-based segmentation and guided region growing for the automatic detection of three PV patterns on the right side of the LA in CT images [10]. This approach, however, is unsuitable for MRI because of image noise, lack of gray-level calibration, field inhomogeneities and artifacts due to patient movement. Incorporating anatomical variations into model-based segmentation remains a challenging and active area of research.

In our proposed approach, model-based segmentation as described in [11] is carried out with multiple models. We define a measure of the local segmentation fit in the areas where the models reflect the anatomical variations without the availability of ground truth annotations. This measure is then used as the input for a support vector machine (SVM) to automatically select the model resulting in the most accurate segmentation.

2   Method

2.1   Used Image Database

The used database consists of 59 whole heart (WH) scans of individual patients which were classified according to the anatomical variant by a clinical expert. About 53 % of the atria had normal anatomy (typical or short CLT on the left and two PVs on the right), 14 % the CLT and 25 % the RMPV pattern. The scans were acquired on Philips Achieva 1.5T systems at Kings College London using either contrast enhanced inversion recovery turbo field echo (IR-TFE) or steady state free precession (SSFP) protocols, both with SENSE encoding for shorter acquisition times. The same cardiac and breathing cycle was achieved using a navigator technique. The image resolution ranged from 0.72 to 1.48 mm in-plane and from 1.5 to 2.0 mm through-plane.

2.2   Shape Models of the Left Atrium

To create models for the most frequent anatomical variations, 32 WH scans were arbitrarily selected from the database: seven with typical anatomy, eleven with short CLT, seven with CLT and seven with RMPV pattern. We used the left atrium model described in [11] as the basis for the model generation. This model has two PVs on each side and was complemented by the trunk of the left atrial appendage (LAA). The model was manually adapted to one arbitrary dataset and an initial feature training was carried out to allow for coarse automatic segmentations of further datasets, which were manually refined to generate the ground truth. Training was carried out on 18 datasets, having either typical anatomy or short CLT on the left and two PVs on the right to represent the normal anatomy (*normal model*). To generate a model for the most frequent anatomical variation on the left side of the LA, the CLT, selective training on seven CLT datasets was performed (*CLT model*). For the RMPV model, a generic cylinder was manually merged to the original mesh, estimating its position, to account for the accessory PV joining the LA. Using this mesh in ground truth annotation for the RMPV-datasets and the following training led to the *RMPV model*. The resulting mean meshes for all three models are depicted in Fig. 1.

(a) *normal model*   (b) *CLT model*   (c) *RMPV model*
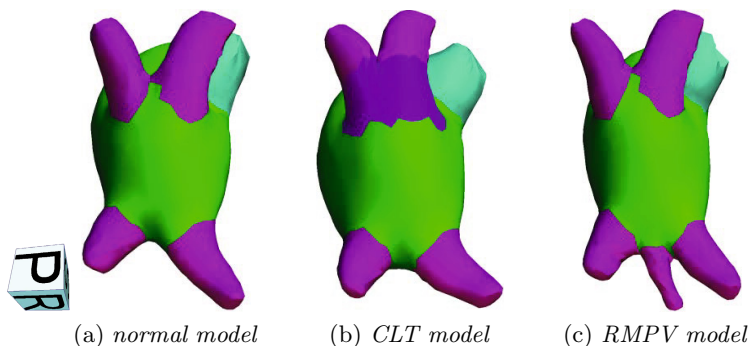
Fig. 1. Resulting labeled mean meshes after training. The left atrial body is labeled green. All pulmonary veins are colored in violet. The LAA is colored in mint green. A special label was assigned to the CLT in dark violet (b).

2.3   Segmentation Framework

In order to generate a patient specific segmentation, consisting of a triangle mesh with labeled triangles, the automatic model-based segmentation framework described in [11] was used. In a nutshell, a hierarchical model of the organ with boundary detection functions trained on reference data is adapted to the 3-D image data. The attraction of the model to the image boundaries is realized by the search for target points, which is carried out for every triangle center along discrete positions on a search profile in the direction of the triangle normal. Target points are selected at the position having the largest weighted feature

response, reflecting the reliability of the found image feature (eg. gray-level or gradient). The aim of the following adaptation steps is to successively minimize the distance between triangle centers and target points. In the beginning, the PVs and the LAA are inactive and only the LA body is adapted to the image. By applying a *single similarity transform*, the pose is corrected. Further adaptation is achieved by employing an *affine transform* to the whole mesh. Next, a *multi-affine adaption* is performed. To enhance segmentation accuracy, we assigned an extra transformation to the right side of the LA. In the last segmentation step a *deformable adaption* is performed to allow the triangle mesh to fully adapt to the patient's anatomy. In this step, the LAA and the PVs are successively activated and adapted to the image. Throughout the entire segmentation process, an external energy term is used that represents the image forces pulling the mesh towards image features. During deformable segmentation, the shape is constrained by an internal energy term which maintains the similarity to the original mean mesh.

For evaluation purposes, leave-one-out crossvalidation on the ground truth data was employed. Segmentation with the four chamber heart model, described in [11], previously trained on MRI, has been used to initialize the position of the LA. Symmetric constrained point-to-surface (CPS) distances between the automatically adapted meshes and the ground truth segmentations were computed according to [11]. The mean CPS distances ϵ_{mean} are shown in Tab. 1. The overall CPS distance of about 1.1 mm for all models is in the range of the voxel resolution. Towards the more variable structures, like the LAA and the PVs, the segmentation error rises but averages to less than 2 mm. Compared to Ecabert et al. [11], who used CT-data with high resolution, the mean errors are larger. This difference is mainly accounted to the coarser resolution of the image data used in our work and the more challenging image characteristics of MRI.

Table 1. Mean constrained point-to-surface distances ϵ_{mean} [mm] resulting from the leave-one-out crossvalidation for the three models

Model	ϵ_{mean} per region [mm]								
	LA body	LSPV	LIPV	CLT	RSPV	RIPV	RMPV	LAA	Whole mesh
Normal	0.87	1.56	1.81	–	1.41	1.11	–	1.70	1.14
CLT	0.81	1.17	1.73	1.18	1.17	1.16	–	1.25	1.01
RMPV	0.79	0.94	0.78	–	1.27	1.60	1.00	1.51	1.05

2.4   Automatic Model Discrimination

To improve the model-based segmentation the three models have to be automatically selected for the respective datasets. The developed models possess differences only in certain regions, leading to different segmentation results in those regions on the same dataset. Information about the segmentation fit is extracted in these regions of interest (ROIs) (see Fig. 2) and used as input for the SVM.

The classification problem is split into two binary classification sub-problems, testing each *variant* against the *normal model*. In the **first sub-problem** the left sides of the two models that are expected to result in different outcomes in the region of the left PVs are considered. Thus, the ROIs of the *nomal model* and the *CLT model* are chosen (Fig. 2a and 2b). Using the analogous argumentation for the right side of the models, the ROIs in the *normal model* and the *RMPV model*, highlighted in Fig. 2c and 2d, are selected for the **second sub-problem**.

(a) *normal model*   (b) *CLT model*   (c) *normal model*   (d) *RMPV model*
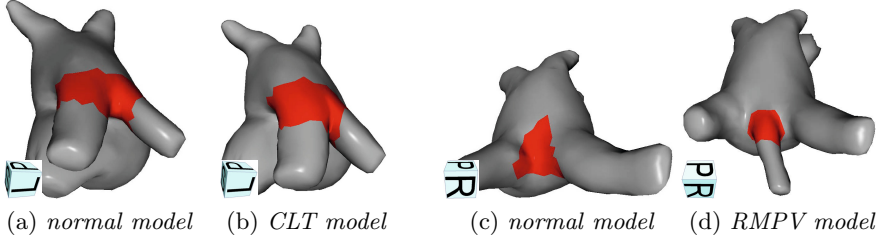
Fig. 2. Regions of interest (ROIs) highlighted in red. For the first sub-problem, the regions in (a) and (b) are considered. For the second sub-problem, the regions in (c) and (d) are selected.

Assessment of the segmentation fit exploits the fact that the model is only approximately adapted to the detected target points. To assess the segmentation fit, the search for target points is started again after the segmentation process is complete to get the distance to the target point and the corresponding feature response for each triangle center. We considered two approaches to construct the observation vectors for the SVM. Firstly, the vectors were constructed using spatial coding, ordered after the triangle index. Secondly, the ROIs were statistically analyzed and the resulting histogram bins were used as the elements of the observation vector. Considering different search profile lengths (2, 4 and 10 mm), we extracted 24 different kinds of SVM input vectors as candidates for our discriminative measure. As input data for the SVM, the histograms of the unsigned distances to the target point at a search profile length of 2 mm were found to perform the best.

For the training of the SVM, the complete image database of 59 WH scans was considered to raise the number of available data and at the same time reducing the bias introduced on the datasets already used in the training of the models. All datasets were automatically segmented with the three models and selected as training data if one of the models yielded a very good segmentation result. For the annotated 32 datasets a mean CPS distance in the area of all ostia of less than 2 mm was considered to be very good. The automatic segmentation results for the remaining datasets were qualitatively evaluated and required to be similarly accurate in the regions of the ostia. In total, 42 datasets, of which 23 have been used in ground truth annotation, were selected for training of the SVM. To optimize the kernel-parameters for the Gaussian kernel, we performed the grid search approach with 5-fold crossvalidation according to [12] and adjusted the class weights to reflect the distribution of the PV-patterns in the image data.

3   Results

3.1   Discrimination Performance for Accurate Segmentations

The aim of the first experiment was to evaluate the automatic model discrimination under optimal conditions. Leave-one-out crossvalidation was performed on the 42 very good training datasets. The class distribution of the datasets is as follows: 57 % normal anatomy, 31 % RMPV, 12 % CLT pattern. For the **first sub-problem** in 40 of the 42 cases the model according to the clinical classification was selected. For the **second sub-problem** the correct model was chosen for 39 datasets (92.86 %). Combining the two sub-problems, the correct model was chosen for 88.1 % of the datasets, meaning that only five of 42 datasets were falsely classified. The confusion matrix summarizes the classification results (Tab. 2). Examination of the automatically selected segmentations of the five datasets revealed that in one case, the automatically selected model (*CLT model*, yellow in Fig. 3a) yielded an absolutely comparable segmentation result, despite not matching the clinical classification (short CLT), as indicated by brackets in Tab. 2. Three datasets with a present third PV were not correctly classified by the SVM. In two of the cases the third PV has an early branching in the ROI of the RMPV model. In the other misclassified case the accessory PV has very low contrast. Taking into account that in one case the model chosen by the SVM yields a similar result, despite not matching the clinical classification, an accurate automatic segmentation is achieved for 90.5 % of the datasets which means an increase of 33.5 % compared to using a single model covering the majority class.

3.2   Discrimination Performance for Average Segmentations

We evaluated the model discrimination method in a second experiment with the 17 remaining datasets which were previously excluded because the automatic segmentation did not fulfill the criterion used to identify very accurate segmentations. It is known a priori that five of the datasets cannot be segmented satisfyingly with either of the models because their variants are not considered. Seven of the twelve remaining datasets can be segmented using the model of the majority class. In nine of the twelve cases, the SVM chose the model corresponding to the clinical classification for the **first sub-problem**, while only one dataset was falsely classified in the **second sub-problem**. The confusion matrix in Tab. 3 includes the results of the previous experiment to reflect the performance on average data. Summing it up, for the complete database, the model corresponding to the clinical classification was automatically chosen in 78.0 % of the cases. As in the first experiment, we further examined the segmentation results of the misclassified cases. With the automatic model selection approach one dataset with a normal PV configuration was classified in the CLT class which resulted in an more accurate segmentation of the left PVs (see Fig. 3b). The RMPV was detected in one of two cases while simultaneously classifying the left side as CLT which yields a more accurate segmentation. In the other case the accessory PV is surrounded by structures that are not visible in the other datasets.
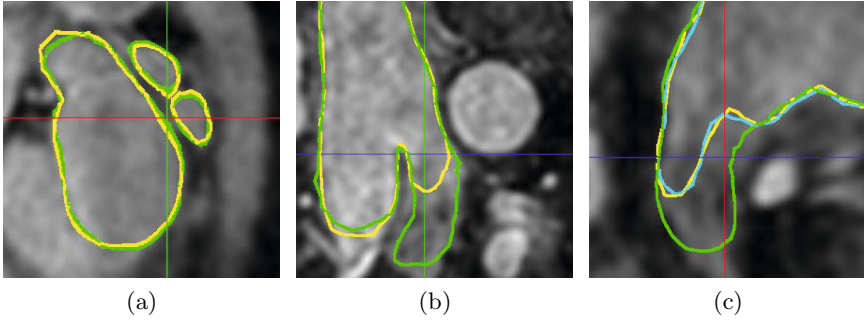
Fig. 3. Examples for successful selection of the better fitting model. Automatic segmentation results with the normal model is shown in yellow, model CLT is shown in green and model RMPV in blue.

Table 2. Confusion matrix for accurate segmentations. Values in brackets indicate false automatic classifications yielding an increased segmentation accuracy.

Actual	Prediction			
Class	normal	CLT	RMPV	CLT+RMPV
normal	23	(1)	0	0
CLT	1	4	0	0
RMPV	3	0	10	0
other	–	–	–	–

Table 3. Confusion matrix for average segmentations

Prediction			
normal	CLT	RMPV	CLT+RMPV
29	(2)	0	0
2	6	0	0
4	0	10	(1)
2	1	2	0

Especially in the region of the ROI, model RMPV does not yield an accurate segmentation. Considering the datasets that were classified falsely but result in a more accurate segmentation with the automatically selected model, 10 of the 12 additional datasets were segmented correctly. For the complete database an accurate segmentation has been achieved in 28.8 % more of the cases compared to using only one model covering the majority class.

4 Conclusion

We present a new approach that enables multi-model-based segmentation by comparing the local fit of different adapted models and automatically choosing the best model using an SVM. Using the three created models, the method achieved a correct clinical classification in 88.1 % of the cases and increased the fraction of datasets that could be accurately segmented from 57 % to 90.5 % under ideal conditions. If accurate segmentations are not achieved, the classifier performs less reliably as shown in the second experiment. Further improvements could be achieved by the generation of new models covering more variants.

References

1. Fuster, V., et al.: ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation–executive summary. Circulation 114(7), 700–752 (2006)
2. Haïssaguerre, M., et al.: Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins. N. Engl. J. Med. 339(10), 659–666 (1998)
3. Marom, E.M., et al.: Variations in pulmonary venous drainage to the left atrium: Implications for radiofrequency ablation. Radiology 230(3), 824–829 (2004)
4. Kaseno, K., et al.: Prevalence and characterization of pulmonary vein variants in patients with atrial fibrillation determined using 3-dimensional computed tomography. The American Journal of Cardiology 101(11), 1638–1642 (2008)
5. Ecabert, O., et al.: Automatic model-based segmentation of the heart in CT images. IEEE Trans. Med. Imag. 27(9), 1189–1201 (2008)
6. Peters, J., et al.: Optimizing boundary detection via simulated search with applications to multi-modal heart segmentation. Med. Image Anal. 14, 70–84 (2010)
7. Zheng, Y., et al.: Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imag. 27(11), 1668–1681 (2008)
8. Manzke, R., et al.: Automatic segmentation of rotational X-ray images for anatomic intra-procedural surface generation in atrial fibrillation ablation procedures. IEEE Trans. Med. Imag. 29(2), 260–272 (2010)
9. Zheng, Y., Wang, T., John, M., Zhou, S.K., Boese, J., Comaniciu, D.: Multi-part Left Atrium Modeling and Segmentation in C-Arm CT Volumes for Atrial Fibrillation Ablation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 487–495. Springer, Heidelberg (2011)
10. Hanna, R., Barschdorf, H., Klinder, T., Weber, F.M., Krueger, M.W., Dössel, O., Lorenz, C.: A Hybrid Method for Automatic Anatomical Variant Detection and Segmentation. In: Metaxas, D.N., Axel, L. (eds.) FIMH 2011. LNCS, vol. 6666, pp. 333–340. Springer, Heidelberg (2011)
11. Ecabert, O., et al.: Segmentation of the heart and great vessels in CT images using a model-based adaption framework. Med. Image Anal. 15, 863–876 (2011)
12. Hsu, C.W., et al.: A practical guide to support vector classification. Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan (2003)

Curvilinear Structure Enhancement with the Polygonal Path Image - Application to Guide-Wire Segmentation in X-Ray Fluoroscopy

Vincent Bismuth[1,2], Régis Vaillant[2], Hugues Talbot[1], and Laurent Najman[1]

[1] Université Paris est, laboratoire d'informatique Gaspard-Monge, équipe A3SI, ESIEE, 77454 Marne-la-Vallée Cedex 2, France
`vincent.bismuth@med.ge.com`
[2] General Electric Healthcare, 283 rue de la minière, 78533 Buc, France

Abstract. Curvilinear structures are common in medical imaging, which typically require dedicated processing techniques. We present a new structure to process these, that we call the polygonal path image, denoted $\mathfrak{P}$. We derive from $\mathfrak{P}$ some curvilinear structure enhancement and analysis algorithms. We show that $\mathfrak{P}$ has some interesting properties: it generalizes several concepts found in other methods; it makes it possible to control the smoothness and length of the structures under study; and it can be computed efficiently. We estimate quantitatively its performance in the context of interventional cardiology for the detection of guide-wires in X-ray images. We show that $\mathfrak{P}$ is particularly well suited for this task where it appears to outperform previous state of the art techniques.

Keywords: curvilinear structures, image segmentation, shortest path.

1 Introduction

Curvilinear structures appear naturally in the human body and thus in medical images. Their segmentation is a specific but extensively studied topic that covers a wide variety of naturally elongated biological structures and medical tools: vascular and cerebral structures and interventional tools like catheters, guide-wires, etc. In this paper, we propose a locally shortest path technique for the processing of curvilinear structures. The proposed technique associates a generalization of several state of the art techniques, an intuitive parameterization, and an efficient computational scheme. We illustrate its performance for the difficult task of guide-wire segmentation in X-ray fluoroscopy. We propose quantified results on clinical data and comparison with other state of the art techniques.

2 Background

2.1 Guide-Wire Detection in X-Ray Fluoroscopy

Interventional radiology/cardiology therapies imply inserting guide-wires into the vascular system of patients under the monitoring of X-ray video, called fluoroscopy. Such procedures are minimally invasive and have been used increasingly

often in recent years, in areas ranging from coronary angioplasty to tumor embolization. Over the past years, guide-wire detection in fluoroscopic images has gained interest and maturity among the image processing community [1,2,3,4]. A large number of applications rely upon its characteristics, such as visualization enhancement, 3D guide-wire reconstruction and respiratory motion tracking. In X-ray images, guide-wires appear as thin, dark curves (see Fig. 1 (a)). The challenge in their detection arises from their low contrast to noise ratio and the superposition with disturbing clutter and anatomical structures.

2.2   State of the Art

Curvilinear structure segmentation techniques and particularly guide-wire detection in X-ray fluoroscopy is often presented as a 3-step pipeline: (1) The **local** building of a feature map representing the probability of presence of an elongated structure at each pixel [2,3,4,5,6]. Its computation often involves considering the neighborhood of each pixel, on which structures are assumed to be straight. (2) A **semi-local** feature map reinforcement, typically enhancing responses aligned along a pattern - e.g. 2nd order, circle or parabola [1], or tensor voting and coherence enhancing diffusion. (3) A **global** structure segmentation. At this stage, simple operators like thresholding are generally not sufficient, and a higher level process is invoked to segment the feature map. It may involve grouping [3,4] or tracking [6].

However some techniques, like locally shortest paths [7] and geodesic path voting [8] are exceptions to this framework. They discard complex curve models (linear, 2nd order etc.) instead relying on the concept of path. Locally shortest paths associate a locally optimal path to every pixel and use the cost of such paths for segmentation. Geodesic path voting relies on a starting point on the structure to segment from which a set of shortest paths are computed to a large number of automatically determined endpoints in the image. The geodesic density (number of paths passing through a pixel) is used to enhance lines.

3   Method

3.1   The Polygonal Path Image

Presentation: As noted above, most line enhancement techniques select at each pixel a curve that best fits locally following a model: line segment, 2nd order model or "arbitrary" smooth curves. We propose a single tunable model that generalizes these steps. Our local curve model is a path, that is characterized by length, and smoothness. Our aim is still to *select at each pixel a best-fitting curve of given smoothness and length*. However, instead of assigning to each pixel a measure associated to a local curve fitting, we propose to use the whole set of locally fitted curves over the image to derive more comprehensive line enhancement techniques.

The smoothness of the paths is controlled by considering regular polygonal paths defined by two parameters: a total arclength L and a length l of every

line segment. These two parameters have a simple intuitive meaning and are set according to the properties of the structure to enhance. The polygonal constraint is a generalization of the classical local path approach [7] (for $l = 1$) enabling the encoding of the a-priori tortuosity of the structures to detect. From a higher level perspective, L is equivalent to the steps of the previous section. Indeed a single segment in the polygon is equivalent to the local step (i.e. segment matching). Considering a few segments is like semi-local processing, reinforcing the aligned responses of the local step. Finally a large number of segments takes into account long paths as in the global step.

Definition of $\mathfrak{P}$: We consider a potential image I of strictly positive values, with structures of interest exhibiting lower values. For each pixel $p \in I$, Φ_p is a set of admissible paths. We assign to each such path $\mathcal{P} \in \Phi_p$ a cost $J(\mathcal{P})$ that is the sum of the values of I along $\mathcal{P}$ (Eq. 1). The path yielding the lowest cost in Φ_p is denoted the locally optimal path $\mathcal{P}_p^*$ at pixel p (Eq. 2). The set of admissible paths is defined by some constraints: Paths shall be regular polygonal lines of given length L, and shall have controlled curvature: at each pixel the path shall be included inside a cone of given orientation and aperture [7]. These constraints enforce locality, the smoothness that is often expected in medical images, and basically defines a search range around each pixel. We call the "polygonal path image" (noted $\mathfrak{P}$) the structure that contains a path and its cost for each pixel.

$$J(\mathcal{P}) = \sum_{p \in I} I(p).\mathbf{1}_{\mathcal{P}}(p) \quad (1) \qquad \qquad \mathcal{P}_p^* = \operatorname*{argmin}_{\mathcal{P} \in \Phi_p^L} J(\mathcal{P}) \quad (2)$$

Complexity: The computation of $\mathfrak{P}$ with a brute force approach has exponential complexity with regard to L: $O(Lk^L)$. Computing all the locally optimal paths in the image for n polygonal segments before considering the optimal paths of $n + 1$ segments, allows us to achieve linear complexity $O(L)$ by adapting the original algorithm of Vincent [7]. This makes it possible to consider $\mathfrak{P}$ with long paths (e.g. one hundred pixels). For example, it takes approximately 10s to process a 512^2 image with such long paths. In terms of memory, we require the initial image times the number of polygon segments to store the paths.

3.2   Structure of the Path Image

Some paths originating from random locations on a clinical image are shown on Fig. 1 (b). We note that paths tend to converge into bundles around the main linear structures, where the cost in the potential image is lower. Those originating outside of the linear structures take the shortest possible way to reach them. Those that start directly on linear structures stay on them until the end or until paths constraints are exceeded. High path density is thus characteristic of the linear structures. Also, the set of paths intersecting at a pixel convey some local geometry information on the linear structures. Fig. 1 (c) illustrates paths

(a)	(b)	(c)	(d)
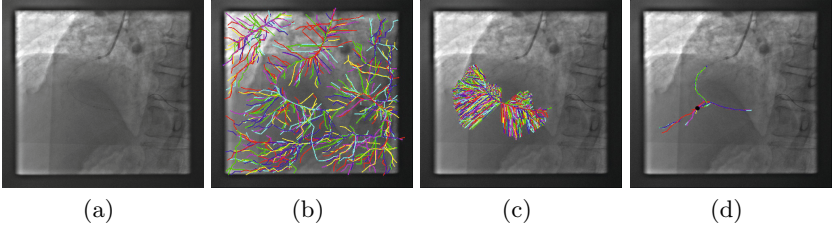
Fig. 1. (a) X-ray fluoroscopy image from an angioplasty exam illustrating a guide-wire, with a long smooth curve appearance and low contrast to noise ratio. (b): 500 locally optimal paths originating from random locations. Observe their tendency to converge to the linear structures of the image and especially to the guide-wire. (c,d) The set of paths intersecting at one given point (belonging to the guide-wire, in (c), and to the background in (d), in this case the point is indicated by the dark spot).

intersecting at a point on the guide-wire. We observe (i) a high number of paths crossing at this point – nearly 11 000 in this case (ii) that they are aligned with the guide-wire in a small neighborhood around the considered pixel.

3.3   Image Processing with $\mathfrak{P}$

Path Voting: We observed that paths form bundles around the main linear structures. Therefore we propose that each minimal path vote for all the pixels it passes through, to reinforce path overlapping. This approach is similar to that of Carlotto [9] and to geodesic voting [8], which also perform voting on sets of paths. Several voting schemes are possible: e.g. each vote can count the same, which is equivalent to counting the number of paths intersecting at one pixel (we note the resulting image $\vartheta(\mathfrak{P})$ - Eq. 3), or each vote can be weighted (with a value denoted $\vartheta_w(\mathfrak{P})$ - Eq. 4). We can also restrict voting to the paths that meet some *a priori* criteria, e.g. preventing very tortuous paths from voting. Such paths occur in parts of the image were there is little relevant linear structure and close to high contrast objects. To this end, we define a tortuosity metric τ to penalize abrupt changes in direction, i.e. 0 for a path with a change of $\pi/2$ rad and 1 for a perfectly straight path. $\tau(\mathcal{P})$ is given by Eq. 5 (where $\mathbf{V}(k)$ is the vector formed by two consecutive vertexes of $\mathcal{P}$). The operation of voting with a path smoothness constraint is denoted $\vartheta_{\tau_{\min}}(\mathfrak{P})$. Its result is illustrated in Fig. 2 (b).

$$\vartheta(\mathfrak{P}) = \sum_{p \in I} \mathbf{1}_{\mathcal{P}_p} \quad (3) \qquad\qquad \vartheta_w(\mathfrak{P}) = -\sum_{p \in I} w(p)\mathbf{1}_{\mathcal{P}_p} \quad (4)$$

$$\tau(\mathcal{P}) = \frac{1}{l^{2(n-2)}} \prod_{k=1}^{n-2} \mathbf{V}(k).\mathbf{V}(k+1) \quad (5) \qquad \vartheta_{\tau_{\min}}(\mathfrak{P}) = \sum_{\substack{p \in I \\ \tau(\mathcal{P}_p) > \tau_{\min}}} \mathbf{1}_{\mathcal{P}_p} \quad (6)$$

Pruning in $\mathfrak{P}$: Another way exploiting the structure of the path image is to prune paths so as to select only a small set of relevant and non-redundant locally

(a)   (b)   (c)   (d)   (e)
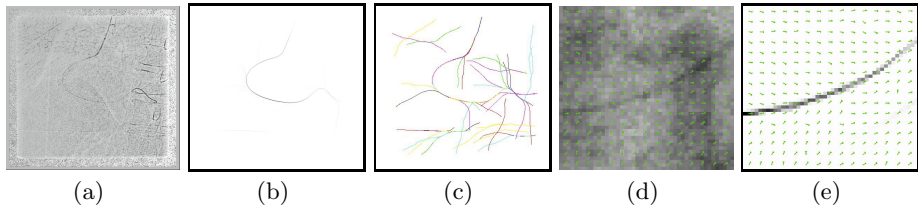
Fig. 2. (a) Dark top-hat of Fig. 1 *(a)*, used as potential image. (b) Result of $\vartheta_{\tau_{\min}=0.6}$. Despite a low CNR the guide-wire is significantly enhanced. (c) Path pruning with $x = 50\%$ and $\tau_{\min} = 0.75$. Observe how the conjunction of pruning and constrained path smoothness help segment the guide-wire. (d) Close-up on a guide-wire segment in Fig. 1 (a) overlaid with the directions extracted from $\mathfrak{P}$. They indicate the direction of the closest curvilinear structure. On the structure itself the direction of the arrow is not relevant, but the orientation is accurate. (e) Same direction field overlaid on $\vartheta_{\tau_{\min}}$.

optimal paths. To do so, we define a neighborhood relationship between paths based on the partial Hausdorff distance. Then, we select the path of minimal cost in $\mathfrak{P}$ and prune all the neighboring paths. We iterate this process in a greedy fashion, selecting the path of minimal cost again and pruning its neighbors, until there is no more path in $\mathfrak{P}$ or some stopping criterion is reached. The remaining set of paths (typically less than 100) describes the linear features in the image (see Fig. 2 (c)).

Estimating Local Orientation: We have illustrated in Fig. 1 that, when they reach a linear structure, paths tend to follow and align with it. Therefore the set of paths intersecting at a pixel on a curvilinear structure can be used to derive its orientation (an angle in $[0, \pi[$). Outside of the curvilinear structures the paths follow the shortest path leading to a linear structure. Their direction thus indicates the shortest path to line-like structures. To estimate it, let us consider at a given pixel p, the set of the paths $\{\mathcal{P}_i\}$ intersecting at p. For each of these paths, we compute the tangent unit vector at point p: $\{\mathbf{t}_i\}$. Then we find the unit vector $\mathbf{v}_p$ yielding the maximal sum of scalar products with the $\{\mathbf{t}_i\}$. This method is illustrated in Fig. 2 (d, e).

3.4   Evaluating and Comparing Line Enhancement Methods

To quantify the performance of our line enhancement techniques, we use ROC analysis to characterize their ability to assign larger values to the pixels of the structure of interest rather than to the background pixels. The Area Under the ROC Curve (AUC) is an estimate of the probability that a classifier will rank a randomly chosen pixel of the structure higher than a randomly chosen pixel of the background. When the number of background pixels far exceeds the number of pixels of the structure of interest, the computation of the AUC can be restricted to the low false positive rate range, in our case $[0, 5\%]$. We normalize the AUC by the range, i.e. 5% here. In this context a random guess

would yield a performance of only 2.5% and a perfect detector 100%. This metric is denoted the "partial AUC". In order to illustrate further the performance of the method we compute the false positive rate for a fixed true positive rate of 90%. Obviously, the better the method, the lower this metric. To compute these metrics, a trained operator manually defined ground truth $\mathcal{T}$ for each image, representing the centerline of the structure. We denote by $\mathcal{D}$ the set of points detected in an image for a given threshold. In order to take into account the real width of the curvilinear structure, we define true positives as the pixels of $\mathcal{T}$ that lie closer to $\mathcal{D}$ than a pre-defined distance d (typically the radius of the guide-wire). False positives are defined as the detected pixels that lie further from $\mathcal{T}$ than d. To make our results unbiased w.r.t. the line-width parameter of the line enhancement techniques, we skeletonize the detected pixels before computing the detection rates. Finally, in order not to corrupt our assessment of line enhancement techniques with the presence of other interventional tools or similar curvilinear structures we also compute the metrics associated to false positives in an ROI around the marked truth.

4   Results

Our clinical case database consists of 12 clinical sequences of 9 images each, for a total of 108 images. These images depict angioplasty exams with the injection catheter, the guide-wire, the angioplasty balloon, the anatomical background, and occasionally, stents and sternal wires. We compared our technique to two state of the art line enhancement techniques : Frangi's Vesselness [5] and Rotated Filter Bank (RFB) [2]. They were found in several recent publications [2,10] to constitute the current state of the art for methods that compute in each pixel the probability to lie in a linear structure. Note that we are aware of dedicated methods for guide-wire segmentation that return a high level object describing the guide-wire [3,4]. However, these methods typically rely on low-level descriptors, and so can benefit from the type of work we present here. We hand-optimized the set of parameters for each technique independently. Regarding Vesselness, we set α and β to the values proposed in Frangi's article [5]. The scale factor σ was set to 2 pixel: the approximate guide-wire radius. For the RFB we relied on a study [10] that concluded that a length of 61 and a width of 3 pixels was optimal for this application. For $\vartheta(\mathfrak{P})$ and $\vartheta_{\tau_{\min}}(\mathfrak{P})$, we set : $l = 21$, $L = 210$, $\tau_{\min} = 0.6$. The potential image on which the path costs are computed was obtained with a simple morphological dark top-hat.

We computed the mean (AUC_μ) and standard deviation (AUC_σ) of the AUC over the database and the FPR for a given TPR of 90%. We report them graphically in Fig. 3. We observe that the line enhancement techniques derived from $\mathfrak{P}$ performed significantly better. This is exemplified by greater AUR values and lower FPR. The local descriptors RFB and Vesselness performed similarly, within a few percents whereas $\mathfrak{P}$ achieves more than twice their performance. In order to illustrate the performance of the different techniques, we selected an optimal threshold that minimizes the sum of the missed detection rate and the

false detection rate. We illustrate the thresholded images obtained with the four methods in Fig. 3. The outcomes are well aligned with the conclusions driven from the partial AUC: $\mathfrak{P}$ based methods perform significantly better. We observe that the results are generally speaking quite satisfactory: most of the guide-wire is segmented, and only a few false positives remain in the background. We can observe how the smoothness constraints in $\vartheta_{\tau_{min}}(\mathfrak{P})$ removes some anatomical false positives that yield tortuous paths. However two problems remain: Over the whole database, the guide-wire tip, which is very contrasted attracts too many paths (creating false positives in its vicinity) and is too tortuous to be accurately fitted by our detector with the same setting as the guide-wire body - It must be detected separately. Secondly, some linear structures are detected in the background that are not guide-wires. Since they satisfy all the properties that we selected for guide-wires, a higher level processing, based on other criteria can handle them. For instance the presence of a tip is very characteristic of the guide-wire, as well as the motion that animates it.
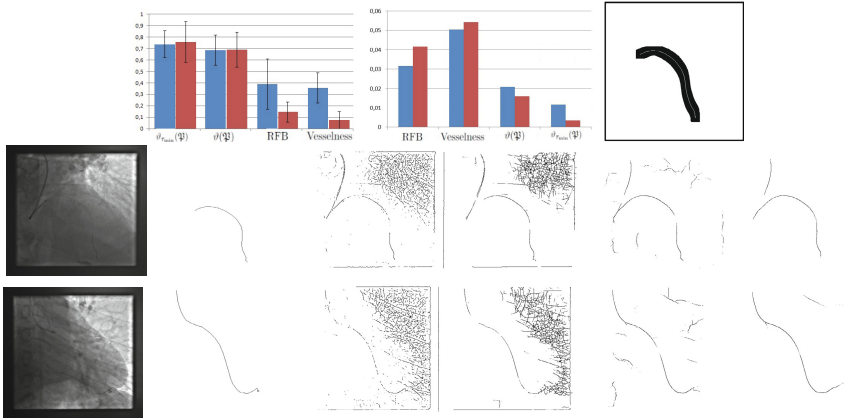
Fig. 3. Top : Performance of the four line enhancement techniques over our database. From left to right: AUC results, the height of the columns is AUC_μ and the error bars represent mean $AUC_\mu \pm AUC_\sigma$. FPR for TPR= 90%. In both graphs the red series is computed over the whole image and the blue one only inside the ROI. The ROI used for FP is a band around the ground truth (see right image). Bottom rows: 2 result examples for optimal thresholds. From left to right : input image, ground truth, Vesselness, RFB, $\vartheta(\mathfrak{P})$ and $\vartheta_{\tau_{min}}(\mathfrak{P})$. Observe that traditional methods enhance a large amount of non relevant structures in the background and fail to enhance some parts of the guide-wire. These problems are not present with $\vartheta_{\tau_{min}}(\mathfrak{P})$. Note these are the exact results of the techniques, no pruning has been performed for the $\mathfrak{P}$ based methods.

5   Conclusion and Further Work

We have presented a new curvilinear structure processing scheme, the polygonal path image. We have demonstrated its suitability for the task of guide-wire

detection in X-ray fluoroscopy. We showed that $\mathfrak{P}$ has several interesting properties: (i) the ability to control the smoothness and length of the structures to be analyzed, (ii) a unification of local, semi local, and global curvilinear structure analysis in a single framework and (iii) an efficient computational scheme. This structure is a rich descriptor of the curvilinear structures present in images from which we derived several tools: line enhancement, segmentation and direction field computation techniques. We have demonstrated the relevance of polygonal path voting schemes for guide-wire segmentation quantitatively in the ROC analysis formalism and compared it to state of the art techniques. Future work may include some natural extensions of the usage/construction of $\mathfrak{P}$, including enhancement and segmentation operators on the path image and the quantitative evaluation of the direction fields and of path pruning. Regarding guide-wire segmentation, $\mathfrak{P}$ is a new approach to the problem that yields significant improvement over state of the art methods. We plan to study in the near future the incorporation of $\mathfrak{P}$ into a complete guide-wire segmentation algorithm.

References

1. Franken, E., Rongen, P., van Almsick, M., ter Haar Romeny, B.M.: Detection of Electrophysiology Catheters in Noisy Fluoroscopy Images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 25–32. Springer, Heidelberg (2006)
2. Bismuth, V., Vancamberg, L., Gorges, S.: A comparison of line enhancement techniques: applications to guide-wire detection and respiratory motion tracking. SPIE Conference Series, vol. 7259 (2009)
3. Barbu, A., Athitsos, V., Georgescu, B., Boehm, S., Durlak, P., Comaniciu, D.: Hierarchical learning of curves application to guidewire localization in fluoroscopy. In: Proc. CVPR, pp. 1–8. IEEE (2007)
4. Honnorat, N., Vaillant, R., Paragios, N.: Guide-Wire Extraction through Perceptual Organization of Local Segments in Fluoroscopic Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 440–448. Springer, Heidelberg (2010)
5. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
6. Meijering, E., Jacob, M., Sarria, J., Steiner, P., Hirling, H., Unser, M.: Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. Cytometry Part A 58(2), 167–176 (2004)
7. Vincent, L.: Minimal path algorithms for the robust detection of linear features in gray images. In: Proc. ISMM, pp. 331–338. Kluwer Acad. (1998)
8. Rouchdy, Y., Cohen, L.: Image segmentation by geodesic voting. application to the extraction of tree structures from confocal microscope images. In: ICPR 2008, pp. 1–5. IEEE (2008)
9. Carlotto, M.J.: Enhancement of low-contrast curvilinear features in imagery. IEEE Transactions on Image Processing 16(1), 221–228 (2007)
10. Tankyevych, O.: Filtering of thin objects: applications to vascular image analysis. PhD thesis, Université Paris Est (2010)

Catheter Tracking via Online Learning for Dynamic Motion Compensation in Transcatheter Aortic Valve Implantation

Peng Wang[1], Yefeng Zheng[1], Matthias John[2], and Dorin Comaniciu[1]

[1] Corporate Research and Technology, Siemens Corporation, Princeton NJ, U.S.A.
[2] Siemens AG, Healthcare Sector, Siemensstr. 1, Forchheim, Germany

Abstract. Dynamic overlay of 3D models onto 2D X-ray images has important applications in image guided interventions. In this paper, we present a novel catheter tracking for motion compensation in the Transcatheter Aortic Valve Implantation(TAVI). To address such challenges as catheter shape and appearance changes, occlusions, and distractions from cluttered backgrounds, we present an adaptive linear discriminant learning method to build a measurement model online to distinguish catheters from background. An analytic solution is developed to effectively and efficiently update the discriminant model and to minimize the classification errors between the tracking object and backgrounds. The online learned discriminant model is further combined with an offline learned detector and robust template matching in a Bayesian tracking framework. Quantitative evaluations demonstrate the advantages of this method over current state-of-the-art tracking methods in tracking catheters for clinical applications.

1 Introduction

Catheter tracking has found important applications in image guided interventions [2,5]. In this work, we present the work of tracking pigtail for dynamic motion compensation in the transcatheter aorta valve implantation (TAVI). A pigtail catheter is a type of medical device that is inserted into the aorta or heart chamber to inject dyes in TAVI. Tracking the motion of a pigtail catheter tip, which is the loose circle at a catheter's distal end, can provide real-time motion information for dynamic model overlay in TAVI [5]. Fig. 1.(a) shows the dynamic overlay of an aorta model (i.e., the solid contour), based on tracked pigtail catheter positions. Compared with the static overlay (i.e., the dotted contour), which does not move with breathing and cardiac motions, the dynamic overlay can update clinicians the position of aorta valves in real-time, therefore could help clinicians achieve high precision in TAVI.

Also shown in Fig. 1. (b)(c)(d), the dynamic and complex clinical environment in TAVI presents great challenges to the catheter tracking. When X-ray images are acquired at arbitrary angles under continuous breathing and cardiac motions, a catheter may show a shape of a circle, an ellipse, or even a straight line in
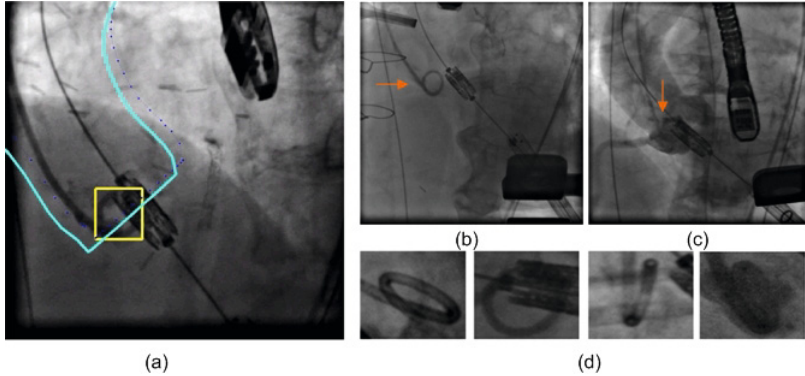
Fig. 1. (a): The dotted contour is a static aorta model projected on 2D, and the solid contour is a dynamically overlayed model based on the tracked catheter (yellow rectangle) ; (b)(c): pigtail catheters in X-ray images; (d): close-up images

a 2D image. Moreover, there also exist in X-ray images many devices, such as other catheters, stents and probes, which could occlude a large part of the pigtail catheter. In past years, there have been work on medical device detection and tracking [2,5]. In [2], an ellipsoid model is used to fit and tracking a lasso catheter. In [5], a SSD (sum of squared differences) based catheter tracking is used. However, only 10 sequences have been tested in the experiments.

To address aforementioned challenges, we present in this paper an online discriminant learning method to adapt the tracking method to dynamic X-ray images. The contribution of this method is that a closed-form analytical solution is developed to efficiently update a discriminant function directly from online images without updating sample co-variance matrices and subspaces as in previous methods [6,8]. The solution is effective and efficient in that it directly minimizes the classification error between tracking objects and backgrounds. Compared with existing online learning work [3,6], our online discriminant learning has a probabilistic formalization that allows us to develop an efficient strategy to adaptively update the discriminant model from online data, instead of selecting from feature candidates [3] or updating scatter matrices and solving eigenvectors as in [6,8]. To handle dynamic changes of pigtail catheters, we further introduce a Bayesian based fusion of multiple measurement models, including the online discriminant model, an offline learned object detection model, and a robust template matching model.

We validate the presented framework on a set of clinical data containing 198 sequences with totally more than 14,000 frames that are captured in several hospitals during cardiac surgeries. Quantitative evaluations and comparison with the state of the art methods [4,8,1] demonstrate that the presented adaptive learning and measurement fusion significantly improve the accuracy and robustness of tracking.

2   Online Discriminant Learning for Catheter Tracking

In this section, we introduce an online discriminant learning method to separate catheters from backgrounds and to handle dynamic environments that could change from one sequence to another. We denote the catheter class as the "positive" class Ω_+ and the background as the "negative" class Ω_-. An observed image is denoted as $\mathbf{Z}$. A data vector extracted from an image patch is denoted as $\mathbf{x}$, and its associated class label is $\ell_\mathbf{x}$. $\ell_\mathbf{x} = 1$ indicates that $\mathbf{x}$ belongs to the catheter class, while $\ell_\mathbf{x} = -1$ indicates that it is from the background.

2.1   Probabilistic Linear Discriminant Analysis

There are many statistical learning methods such as SVM, AdaBoost, k-NN, and Neural Networks. For the consideration of simplicity and computational efficiency, the linear discriminant analysis is used. In the linear discriminant analysis, an original data vector is projected to a subspace of lower dimensionality where objects can be separated from the background. The projection is represented as $\mathbf{y} = \Phi^T \mathbf{x}$, where Φ is the linear transformation vector (or a matrix). Fisher discriminant analysis (FDA) is a commonly used linear discriminant analysis method, which maximizes the ratio of between-class and within-class variance. FDA can be solved as a generalized eigenvector problem:

$$(\Sigma_w^{-1}\Sigma_b)\Phi_k = \lambda_k\Phi_k. \tag{1}$$

where Σ_b and Σ_w are between- and within-class scatter matrices, respectively. The transformation vector Φ_k is the eigenvector corresponding to the k-th eigenvalue. For a two-class problem, only one effective eigenvector can be extracted from FDA. The FDA transformation vector is denoted as $\mathbf{A}_f$. Usually PCA is applied before FDA to reduce the data dimensionality. The overall linear discriminant analysis is the combination of PCA and FDA, as Eqn. (2):

$$y_\mathbf{x} = \mathbf{A}_f^T \mathbf{A}_p^T \mathbf{x} = \mathbf{A}^T \mathbf{x}, \tag{2}$$

where $\mathbf{A}_p$ is the PCA eigenspace projection matrix, and $\mathbf{A} = \mathbf{A}_p\mathbf{A}_f$ is the overall linear discriminant vector. After learning the linear discriminant vector $\mathbf{A}$, a sigmoid function is used to model the posterior probability for the two-class classification:

$$P_\mathbf{A}(\mathbf{x}) = P(\ell_\mathbf{x} = 1|\mathbf{x}) = \frac{1}{1 + e^{(ay_\mathbf{x}+b)}} = \frac{1}{1 + e^{(a\mathbf{A}^T\mathbf{x}+b)}} \tag{3}$$

The parameters a and b can be learned from training samples by regression.

2.2   Online Discriminant Learning

The probabilistic linear discriminant can be learned offline from collected training samples. However, the background, and sometimes the object itself, could

<p style="text-align:center;">(a)   (b)</p>
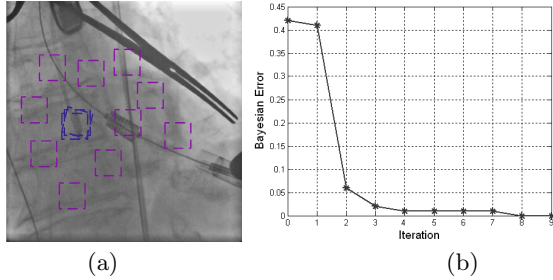
Fig. 2. Online discriminant learning. (a): positive samples (blue rectangles) and negative samples (purple rectangles) are extracted online; (b): the convergence of online learning at a frame.

change from one tracking scenario to another, and may not be seen in training sets, therefore not available for offline learning. In dynamic environments, the discriminant learning needs to adapt to different tracking scenarios. For this purpose, we present an online method to efficiently and effectively update the discriminant vector $\mathbf{A}$.

Most statistical learning algorithm aims at the minimization of the Bayesian error, i.e., $E_f = \int (1 - p(\ell_{\mathbf{x}}|\mathbf{x}))p(\mathbf{x})\mathbf{dx}$. In the adaptive discriminant learning, we also try to minimize the Bayesian error, which is equivalent to minimizing the object function $C(\mathbf{A})$ in Eqn. (4):

$$C(\mathbf{A}) = -[\sum_{\mathbf{x}_i \in \Omega_+} P(\mathbf{x}_i)P_{\mathbf{A}}(\mathbf{x}_i) - \sum_{\mathbf{x}_j \in \Omega_-} P(\mathbf{x}_j)P_{\mathbf{A}}(\mathbf{x}_j)] = \sum_{\mathbf{x}} \frac{-\ell_{\mathbf{x}}P(\mathbf{x})}{1 + e^{(a\mathbf{A}^t\mathbf{x}+b)}}. \quad (4)$$

With new online data available during tracking, to re-train the discriminant analysis needs to access a large number of training samples, and involves recomputation of co-variance metrics and eigenbases. In this method, we online update the discriminant analysis using a gradient descent method. To minimize the Bayesian error, the discriminant analysis is iteratively updated as Eqn. (5):

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k-1)} - \delta_k \nabla_{\mathbf{A}} C(\mathbf{A}^{(k-1)}), \quad (5)$$

where δ_k is an updating step at the k-th iteration. The updating continues until it converges. The gradient of the object function $\nabla_{\mathbf{A}} C(\mathbf{A})$ is given by

$$\nabla_{\mathbf{A}} C(\mathbf{A}) = \sum_{\mathbf{x}_i} \nabla_{\mathbf{A}}\left(\frac{-\ell_i P(\mathbf{x}_i)}{1 + e^{(a\mathbf{A}_i^t\mathbf{x}_i+b)}}\right) = \sum_{\mathbf{x}_i} \ell_i P(\mathbf{x}_i)\frac{ae^{(a\mathbf{A}^t\mathbf{x}_i+b)}}{(1 + e^{(a\mathbf{A}^t\mathbf{x}_i+b)})^2}\mathbf{x}_i. \quad (6)$$

After $\mathbf{A}$ is updated, a and b are also updated by regression at every iteration.

The online discriminant needs a good starting point to converge to an optimal solution, although a global optimum is not guaranteed. A starting point can be the discriminant function learned offline from collected samples. Even if the

initial discriminant vector may not fit a current environment, the online learning can quickly converge to a good solution. An example is shown in Fig. 2, where the greedy searching converges within a few iterations to achieve a Bayesian error less than 10% for the initial error of 40%. To improve the learning robustness, we keep a history of tracked objects from previous frames in a pool to prevent the learning from being distracted by tracking errors at individual frames. The Bayesian error estimated as Eqn. (4) is used as a criterion to decide if the learning is converged to an optimal solution. For computational efficiency, the online updating of discriminant vector can be performed once every several frames.

3   Bayesian Catheter Tracking Framework

The tracking is to infer unknown object states $\mathbf{m}_t$, i.e., the catheter position in this paper, from an observed image sequence $\mathbf{Z}_{1:t}$. We formalize the parameter inference in a sequential Bayesian inference framework. Assuming a commonly used Markov property for tracking, the posterior probability $P(\mathbf{m}_t|\mathbf{Z}_{1:t})$ is given in Eqn. (7).

$$P(\mathbf{m}_t|\mathbf{Z}_{1:t-1}) \ = \int P(\mathbf{m}_t|\mathbf{m}_{t-1})P(\mathbf{m}_{t-1}|\mathbf{Z}_{1:t-1})d\mathbf{m}_{t-1}$$
$$P(\mathbf{m}_t|\mathbf{Z}_{1:t}) \ \propto P(\mathbf{m}_t|\mathbf{Z}_{1:t-1})P(\mathbf{Z}_t|\mathbf{m}_t) \tag{7}$$

The tracking result is the motion parameter corresponding to the maximal posterior probability, i.e, $\hat{\mathbf{m}}_t = \arg\max_{\mathbf{m}_t} P(\mathbf{m}_t|\mathbf{Z}_{1:t})$.

In Eqn. (7), $P(\mathbf{m}_t|\mathbf{m}_{t-1})$ is a dynamic model. We model the dynamic probability as a Gaussian model. The likelihood measurement model $P(\mathbf{Z}_t|\mathbf{m}_t)$ is a fusion of multiple measurements. Assuming that there are K measurement models for an object, i.e., $P^{(k)}(\mathbf{Z}_t|\mathbf{m}_t) = P(\mathbf{Z}_t|\mathbf{m}_t, k), k = 1, ..., K$, a measurement fusion based on the probability marginalization is given in Eqn. (8):

$$P(\mathbf{Z}_t|\mathbf{m}_t) = \sum_k P(\mathbf{Z}_t|\mathbf{m}_t, k)P(k|\mathbf{m}_t) \tag{8}$$

where $P(k|\mathbf{m}_t)$ is the weight for the k-th measurement model.

The online learned discriminant model acts as one measurement model, denoted as $P(\mathbf{Z}_t|\mathbf{m}_t, k = 1) = P_{\mathbf{A}}(\mathbf{x}_{\mathbf{m}_t})$ as in Eqn. (3), where $\mathbf{x}_{\mathbf{m}_t}$ is the data vector extracted from observed images based on the object state $\mathbf{m}_t$. The second measurement model used is an offline learned object detector. A probabilistic boosting tree (PBT) [9] is trained with Haar features to build the catheter detector offline. The probabilistic interpolation of an AdaBoost is used to provide the second measurement model $P(\mathbf{Z}_t|\mathbf{m}_t, k = 2)$. A robust template matching method is used as the third measurement mode in our method. The template matching takes the form in Eqn. (9):

$$P(\mathbf{Z}_t|\mathbf{m}_t, k = 3) \propto G(D(\mathbf{x}_{\mathbf{m}_t}); \sigma_a), \tag{9}$$

where G is a one-dimensional Gaussian kernel with a zero mean and a bandwidth of σ_a. $D(\mathbf{x}_{\mathbf{m}_t})$ is a robust distance function [7] to compute the differences between current observations $\mathbf{x}_{\mathbf{m}_t}$ and a catheter appearance template, which is updated from the tracking result at a previous frame.

4   Experiments

A set of clinical data has been acquired from TAVI procedures in several hospitals to evaluate the catheter tracking. There are totally 198 sequences acquired at the frame rate between 15fps and 30fps, and more than 14,000 frames in the data set. Each image pixel represents a physical size between 0.154 mm and 0.308 mm. The data set well represents the real surgical scenarios, includes images of poor quality, occlusions, dye injections, and motion blurs. Some exemplar frames in the data set and corresponding tracking results are displayed in Fig. 3. To establish ground truth for evaluation, we manually annotate the pigtail catheters in all the frames. The annotation at the first frame of each sequence is used to initialize the tracking, and the annotations at the remaining frames are used for evaluations. The whole system runs at more than 10 frames per second at an Intel Quad Core 2.5GHz CPU.

In our quantitative evaluation, we measure the **tracking precision** as the Euclidean distance between tracked catheters and the ground truth. The **normalized error** is the distance error divided by corresponding pigtail catheter size (the maximum of the width and height of a catheter). The **tracking success rate** is calculated to measure the percentage of successful tracking. The **frame success rate** is defined as the percentage of the frames whose normalized distance error is less than a threshold (e.g., set as 0.35 in our experiments). The **sequence success rate** is defined as the percentage of sequences where more than a certain percentage of frames (e.g., set as 70% in our experiments) have been successfully tracked. For an unbiased evaluation, we perform a **4-fold cross-validation** to evaluate the tracking accuracy. During evaluation, the algorithm parameters are kept the same for all the sequences, so the method is not over-fitted to individual cases.

Some parameters of our method, such as covariance matrices in Gaussian distributions, can be learned. A few other parameters, such as the weights in Eqn.(8), need to be set empirically. Through our experiments, we found that the accuracy of this method is consistent within a range of parameter settings. Due to limited space, only the accuracy of a single set of parameters (i.e., the weights are set as 0.35, 0.25 and 0.4 for the template based, online learned and offline learned measurement models, respectively) is reported in Table 1.

The quantitative evaluation results in Table 1 show that the online models perform better than offline learned detectors, i.e., 84.1% vs. 82.9% of frame success rate and 79.4% vs 73.9% of sequence success rate. When combining all the measurement models together, this method achieves a 91.0% frame success rate and a 91.9% sequence success rate, respectively. The tracking precision, e.g. the mean error of 1.10mm, further confirms the performance improvement with

(a)

(b)

(c)
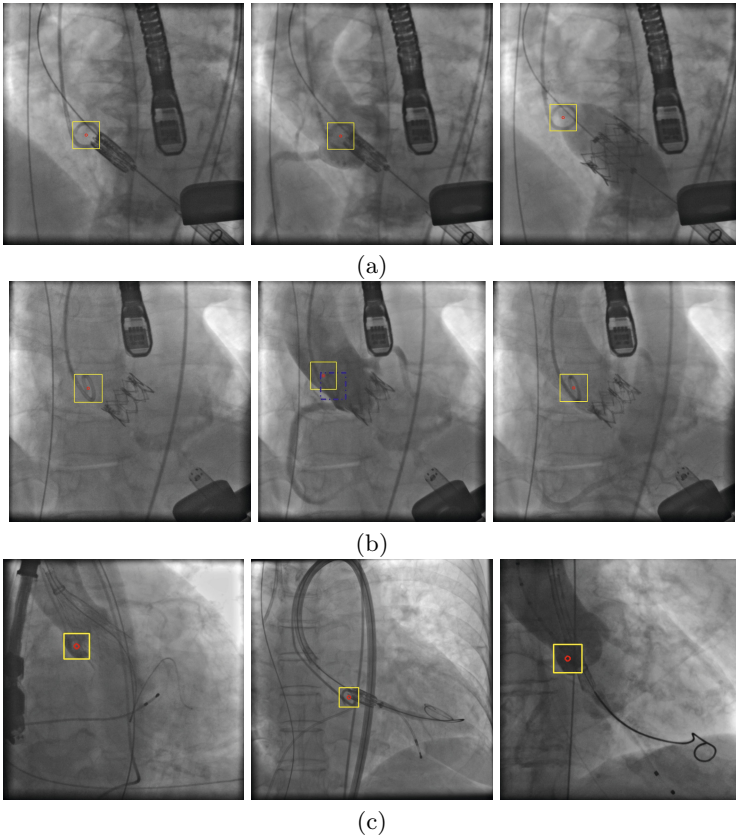
Fig. 3. Exemplar tracking results: (a) handling occlusion and dye injection; (b) recovering from dye injection. The blue rectangle in the middle image shows the ground truth;(c) more tracking results. Each image corresponds to one sequence.

this method. To understand the generalization capability of the method, we also train an object detector with all the annotations and apply it to the tracking. The evaluation, named as "full training", shows that the tracking success rates only drop by 3% with the cross-validation. The tracking failures of our methods are mainly caused by contrast injection and image blurs. In practice, some of tracking failures won't affect the workflow if the tracking resumes correctly after contrast injection, as shown in the example of Fig. 3 (b).

This method is compared with three representative state-of-the-art visual tracking methods: the online boosting method [4], the incremental learning methods [8], and the multiple instant learning method [1], whose implementations are available from authors' homepages. For each algorithm for comparison, we carefully tune the parameters to achieve optimal tracking accuracy. The methods of [4] and [8] can achieve $\sim 65\%$ frame success rate. The multiple instant learning method [1] can only succeed in less than 10% sequences with a mean error of more

Table 1. Quantitative evaluation on 198 sequences

This method	Tracking success rate		Tracking precision (mean errors in mm/normalized errors)	
	Frame	Sequence	median	mean
Using only offline detector(cross-validation)	82.9%	73.9%	1.37/0.08	5.12/0.27
Using only online models	84.1%	79.4%	1.38/0.08	3.23/0.18
Bayesian fusion (cross-validation)	91.0%	91.9%	1.10/0.06	3.19/0.17
Bayesian fusion (full training)	93.3%	94.9%	0.98/0.05	2.43/0.13
Existing methods				
Grabner et. al. [4]	64.4%	52.7%	3.15/0.21	4.55/0.35
Ross et. al. [8]	65.1%	64.3%	2.91/0.17	5.56/0.40

than 10mm. The poor performance of existing methods shows the challenges of pigtail catheter tracking in X-ray images, and demonstrates the advantages of our method in handling such a challenging environment.

5   Conclusion

In summary, this paper presented a novel and robust method to track pigtail catheters for dynamic motion compensation in TAVI. The validation on a set of 198 sequences demonstrated the performance for this tracking task.

References

1. Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: CVPR (2009)
2. Brost, A., Liao, R., Hornegger, J., Strobel, N.: 3-D Respiratory Motion Compensation during EP Procedures by Image-Based 3-D Lasso Catheter Model Generation and Tracking. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 394–401. Springer, Heidelberg (2009)
3. Collins, R., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. IEEE Trans. on PAMI 27(10), 1631–1643 (2005)
4. Grabner, M., Grabner, H., Bischof, H.: Learning features for tracking. In: CVPR (2007)
5. Karar, M.E., John, M., Holzhey, D., Falk, V., Mohr, F.-W., Burgert, O.: Model-Updated Image-Guided Minimally Invasive Off-Pump Transcatheter Aortic Valve Implantation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 275–282. Springer, Heidelberg (2011)
6. Lin, R.S., Yang, M.H., Levinson, S.: Object tracking using incremental Fisher discriminant analysis. In: ICPR, vol. 2, pp. 757–760 (2004)
7. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. IEEE Trans. on Pattern Analysis and Machine Intelligence 26(6), 810–815 (2004)
8. Ross, D., Lim, J., Lin, R.S., Yang., M.H.: Incremental learning for robust visual tracking. International Journal of Computer Vision Special Issue: Learning for Vision (2007)
9. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV, pp. 1589–1596 (2005)

Evaluation of a Real-Time Hybrid Three-Dimensional Echo and X-Ray Imaging System for Guidance of Cardiac Catheterisation Procedures

R.J. Housden[1], A. Arujuna[1,2], Y. Ma[1], N. Nijhof[3], G. Gijsbers[3], R. Bullens[3], M. O'Neill[1,2], M. Cooklin[2], C.A. Rinaldi[2], J. Gill[2], S. Kapetanakis[2], J. Hancock[2], M. Thomas[2], R. Razavi[1,2], and K.S. Rhode[1]

[1] Division of Imaging Sciences and Biomedical Engineering, King's College London, UK
[2] Department of Cardiology, Guy's and St. Thomas' NHS Foundation Trust, London, UK
[3] Philips Healthcare, Interventional X-ray, Best, The Netherlands

Abstract. Minimally invasive cardiac surgery is made possible by image guidance technology. X-ray fluoroscopy provides high contrast images of catheters and devices, whereas 3D ultrasound is better for visualising cardiac anatomy. We present a system in which the two modalities are combined, with a trans-esophageal echo volume registered to and overlaid on an X-ray projection image in real-time. We evaluate the accuracy of the system in terms of both temporal synchronisation errors and overlay registration errors. The temporal synchronisation error was found to be 10% of the typical cardiac cycle length. In 11 clinical data sets, we found an average alignment error of 2.9 mm. We conclude that the accuracy result is very encouraging and sufficient for guiding many types of cardiac interventions. The combined information is clinically useful for placing the echo image in a familiar coordinate system and for more easily identifying catheters in the echo volume.

Keywords: Intervention guidance, image fusion, registration, X-ray fluoroscopy, 4D ultrasound.

1 Introduction

Minimally invasive cardiac interventions are becoming increasingly feasible with developments in image guidance [1]. X-ray fluoroscopy is commonly used for guidance, as it provides real-time images with excellent visualisation of catheters and other devices. However, fluoroscopy provides only a two dimensional (2D) projection image and has poor soft tissue contrast, and so is not sufficient to guide complex procedures. The complementary technology of trans-esophageal echo (TEE) provides real-time volumetric images with good contrast of cardiac anatomy, but is susceptible to artefacts when imaging mechanical devices such

as catheters. Also, TEE has a limited field of view and therefore requires a skilled cardiologist to provide anatomical context. The combination of the information in X-ray fluoroscopy and 4D TEE is potentially useful for procedure guidance.

We have developed a system in which the TEE volume is registered to and overlaid on the X-ray, allowing simultaneous real-time visualisation of cardiac anatomy and interventional devices. This concept has been proposed previously and the various implementations differ mainly in the way the echo volume is registered to the X-ray. Existing approaches include tracking the TEE probe with an electromagnetic or mechanical tracking device [2,3] and tracking the TEE probe head, or markers attached to the probe head, in the X-ray images [4,5]. In this paper, we present a complete real-time clinical system for hybrid echo/X-ray navigation that was developed from the proof-of-concept methodology in [4]. We evaluate the temporal and spatial accuracy of our overlay alignment and present initial clinical experience of our system used in atrial fibrillation (AF) catheter ablation cases and trans-aortic valve implantation (TAVI) procedures.

2   Methods

2.1   System Overview

The scanning setup comprises a Philips Allura Xper FD10 C-arm X-ray system and a Philips iE33 3D ultrasound system with an X7-2t 3D TEE probe. Data is streamed from each system to a PC running the visualisation software and displayed in real-time. The software displays a 2D projection view in X-ray image coordinates, overlaid with a volume rendering of the echo volume.

The two views are registered using the projection image of the TEE probe in the X-ray, following the method described in [4]. A 3D model of the TEE probe head, acquired from a nano-CT scan, is registered to the X-ray. The registration is initialised manually to approximately the correct position and orientation before running an automatic 3D-2D registration algorithm. Following the initial registration, the automatic registration is repeatedly rerun to track changes in the probe position due to cardiac and respiratory motion. The automatic GPU-accelerated registration updates in this way at a rate of 1–2 Hz. Figure 1 shows a typical X-ray view, the probe model and a registration of the two.

2.2   Temporal Synchronisation

The real-time X-ray and echo images are received from different sources and are processed in different ways before being transmitted to the visualisation software. It is important to determine the relative synchronisations of the two data streams in the overlay view. We expect any delay in the X-ray stream to be constant, but the processing applied to the echo volume in the ultrasound machine may depend on various settings, such as imaging depth, which can be changed several times during a clinical procedure.

The temporal synchronisation was determined by imaging a catheter with simultaneous X-ray and echo. The ultrasound machine was set to image in its live
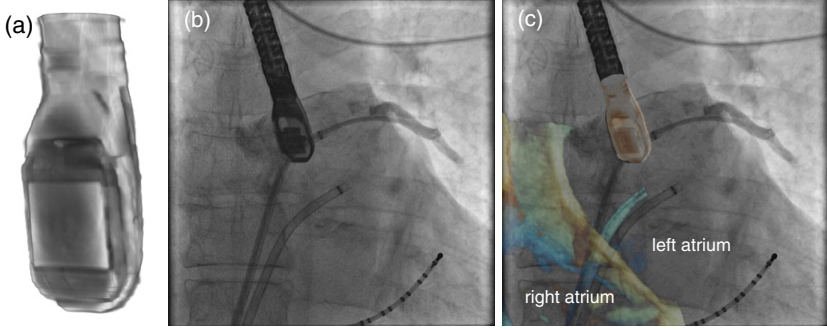
Fig. 1. Automatic overlay registration. (a) TEE probe model from a nano-CT scan of the probe head. (b) X-ray image with the projection of the TEE probe clearly visible. This example is from an AF ablation case. (c) Echo overlay, with the probe model registered to the X-ray by automatic 3D-2D registration.

3D mode. The catheter was given a sharp tap such that the resulting movement was easily detectable in both the X-ray and echo images by sequentially differencing the data. The time difference between the start of movement in the X-ray and echo images provides the relative delay in the overlay view. The delay was measured over the available range of depth settings of the ultrasound scanner (3–26 cm) with 1–3 separate measurements at each depth.

2.3   Spatial Alignment Accuracy

As explained above, the overlay registration is achieved by aligning images of the ultrasound probe. The alignment accuracy of the echo volume to the X-ray image is affected by both the accuracy of this probe tracking and the accuracy of the previously calibrated, fixed transformation from the probe to the echo volume. The alignment is particularly susceptible to orientation errors, as even small angular errors can be extrapolated to large misalignments. These alignment errors were measured in phantom and *in vivo* experiments. The accuracy may also depend on the probe orientation in the projected view. In this paper, the analysis is restricted to probe views typically used in clinical practice.

The alignment error in an overlay view was calculated in terms of the 2D projection error [6]. This error, e, is given by

$$e = \frac{1}{N} \sum_{i=1}^{N} \| P_{i,\mathrm{xray}} - P_{i,\mathrm{echo}} \| \frac{D_{\mathrm{source} \rightarrow \mathrm{target}}}{D_{\mathrm{source} \rightarrow \mathrm{detector}}} \ , \tag{1}$$

where $P_{i,\mathrm{xray}}$ and $P_{i,\mathrm{echo}}$ are the locations of corresponding landmarks defined in the X-ray and echo projections respectively and N is the number of these points

in the image. The distances $D_{\text{source} \rightarrow \text{target}}$ and $D_{\text{source} \rightarrow \text{detector}}$ account for the magnification in the projected image and their values are known from the C-arm geometry, although the value for $D_{\text{source} \rightarrow \text{target}}$ is known only approximately. The error value is an average of the error at several landmarks in the image so that each overlay view gives one error value. The final error is given in mm.

Phantom Experiment. The phantom comprises a water tank in which two wires are suspended forming a cross. This was imaged by the TEE probe with the ultrasound machine in full volume mode. Corresponding X-ray images were acquired at four different C-arm positions (RAO 54, RAO 24, LAO 22 and PA views) each at high and low dose (8 X-rays in total). From this data, alignment errors were measured for two registration strategies. The first approach was automatic registration following an approximate manual initialisation. This was done on three different images in each X-ray sequence giving 24 overlay views in which to measure errors. The second approach was a careful manual alignment. Note that this manual alignment considered only the probe, not the echo volume, when optimising the alignment. Again, this was done for 24 overlay views.

Landmarks were defined by manually fitting straight lines to the crossed wires in the X-ray and echo projection image. The crossing points were detected automatically from these lines and landmarks were automatically defined in fixed steps along the lines from the crossing point. This provided up to 13 landmarks per overlay view (minimum 10 where the crossing point was near the edge of the view). 2D projection errors were calculated between corresponding landmarks and averaged according to (1). Figure 2(a) shows an example of this.
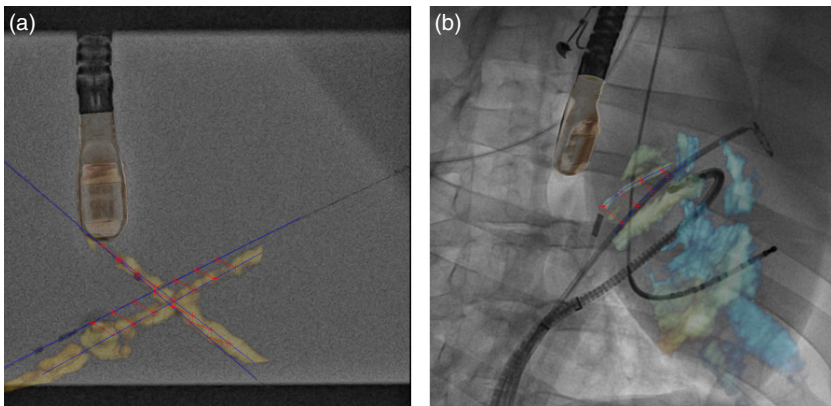
Fig. 2. Error measurement in overlay views. (a) Phantom experiment overlay example. Errors are measured between automatically defined landmarks on straight line models of the crossed wires. (b) Porcine scan overlay example. Errors are measured from landmarks on the echo catheter image to their nearest point on a spline model of the X-ray catheter.

***In Vivo* Accuracy.** *In vivo* data to measure accuracy were acquired from a live porcine experiment. Five echo volumes were acquired in full volume mode with the probe in different positions. At each position, three X-ray sequences were recorded in RAO 30, LAO 30 and PA positions with high X-ray dose setting (15 X-ray views in total). Various catheters were inside the heart during the acquisition and these provided convenient targets for measuring alignment errors. Errors were again measured for automatic and manual registrations. These were done at two separate frames in each X-ray sequence (30 measurements), manually selected from the same cardiac and respiratory phases.

Corresponding catheters were manually defined in the echo and X-ray views using Catmull-Rom spline curves. Equally spaced points along the echo curve were automatically selected as echo landmarks. The corresponding X-ray landmark was defined as the closest point on the X-ray curve. The alignment error for each overlay view was again taken as the average of the 2D errors between landmarks according to (1). An example of these error measurements is given in Fig. 2(b). Average errors were measured using between 3 and 6 landmark pairs per overlay view, depending on the length of catheter visible in the echo image. It should be noted that our approach of measuring to the closest point does not necessarily capture the complete error as there can also be misalignment tangentially to the catheters.

Clinical Cases. The system has been used alongside standard imaging technologies during two types of minimally invasive cardiac procedures: AF ablation and aortic valve replacement (TAVI). During each case, X-ray and echo sequences were recorded independently. Although real-time synchronised visualisation of the live data stream was possible during the procedure, the post-procedure analysis for this paper required that the recorded X-ray and echo data were matched up manually, resulting in only approximately synchronised sequences. 13 such sequences were successfully reconstructed from seven of the nine patients undergoing ablation procedures and six sequences were generated from two TAVI procedures. The errors in each sequence were considered separately.

Unlike the phantom and porcine experiments, the clinical data were analysed in sequences of overlay views over a period of time. There may therefore be additional errors from the real-time probe motion tracking, which may not be able to keep up with sudden or continuous probe movements. Alignment accuracy was measured over the sequences by considering multiple images in each sequence (between 5 and 21 images depending on the sequence length). Statistics for the error over time were calculated for each sequence separately. Errors were calculated using landmarks defined on the catheters and devices in the same way as for the porcine data.

3   Results and Discussion

3.1   Temporal Synchronisation

Figure 3 shows the results of the time delay experiment. In general, the echo and X-ray are synchronised to within 150 ms. The results show a significant change

($p < 0.01$) at around 18 cm depth. At smaller depth settings, the echo lags the X-ray by 46 ± 43 ms. At larger depths, the echo leads by 82 ± 42 ms. This appears to be a sudden change rather than a gradual trend. Therefore, temporal synchronisation could be substantially improved by using a simple pre-calibrated time shift dependent on the depth cut-off value of 18 cm.
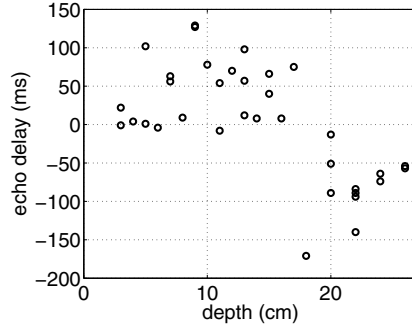
Fig. 3. Relative delay of X-ray and echo. The plot shows the lag of the echo relative to the X-ray over the full range of depth settings.

In terms of clinical implications, the relative delay in the two image streams means that there is a time difference between the display of the catheter locations (most clearly visible in the X-ray image) and the cardiac anatomy (only visible in the echo volume). In a typical cardiac cycle (around 1 s), the time difference is on average less than 10% of the cardiac cycle and at most is within 20%. The visual impact of the delay will depend upon the frame rate of the data streams (maximum 30 frames per second (fps)). For electrophysiology procedures, X-ray frame rate can be as low as 3 fps to minimise radiation dose. The measured delay is then likely to have no impact on ease of navigation using the hybrid approach. The delay is also small compared to the update rate of the 3D-2D registration (1–2 Hz), although this is only significant when the probe moves substantially.

3.2   Overlay Alignment Accuracy

Figure 4 shows the results of the overlay alignment error measurements. Considering first the phantom experiment, the images of the crossed wires are clearly best aligned when positioning the probe manually, which serves as a ground truth. The errors for automatic registration were 6 mm on average. These results included all registrations even though many would be visually deemed as failures. Detection of registration failure, possibly by use of the similarity metric, would substantially reduce these errors and make them approach the ground truth result. Manual alignment in the porcine data shows similar results to the automatic alignment with average errors of approximately 6 mm. Part of this error will be due to the X-ray and echo data not being recorded simultaneously.
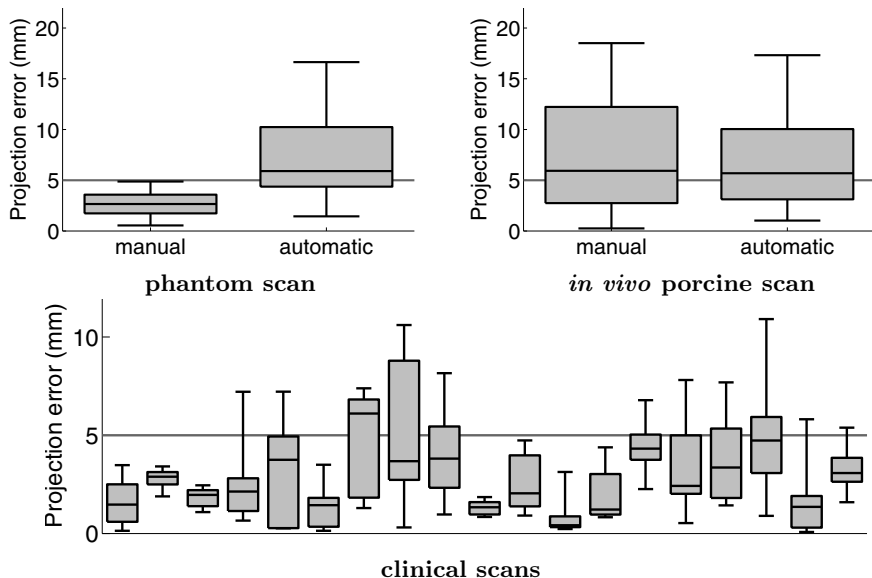
Fig. 4. Overlay alignment errors. The graphs show box plots for the alignment error measurements. Percentiles are at 2.5, 25, 50, 75 and 97.5%. In the clinical results, the first 13 box plots are for sequences taken from the AF ablation cases and the remaining six are from the TAVI cases. Lines drawn at 5 mm error are for reference.

The overall average error of the clinical data is 2.9 mm, with a maximum error of 10.9 mm. The accuracy requirement for a clinically useful image guidance system depends on many factors including the patient and the procedure being performed, as discussed in [7]. In general, many of the errors are well below 5 mm with only a few cases that are worse. Again, detection of registration failures will improve accuracy. The result of 2.9 mm is very encouraging, especially since it represents a measure of the overall system accuracy in the clinical setting.

A typical overlay from a clinical AF case is shown in Figure 1(c). The clinical results show that the system is sufficiently accurate to guide many different types of cardiac interventions. An advantage of the overlay view is that it helps in interpreting low quality echo images. An echo volume on its own can be difficult to interpret because of its limited field of view and lack of context for the echo coordinate system relative to the patient. Also, catheters and devices tend to produce artefacts in the ultrasound data reducing the clarity of the images. In the overlay view, the echo volume is displayed in a coordinate system that can be more easily related to the patient by the experience of cardiologists working routinely with X-ray fluoroscopy. Also, the highly visible catheters in the X-ray image help with identifying the catheters in the echo and so can be related to the cardiac anatomy via the echo image. In this way, a registered overlay view provides useful information for procedure guidance.

4   Conclusions

In this paper, we have presented a platform for real-time hybrid X-ray fluoroscopy and 3D echo visualisation and have successfully demonstrated it in phantom, animal and clinical experiments. The system exhibits a temporal synchronisation that is less than 10% of a cardiac period on average. This will not have a significant impact on guidance at the low fluoroscopy frame rates that are often used to minimise radiation exposure. The average 2D registration error in clinical data is 2.9 mm which is encouraging for the system's eventual use in clinical guidance. The hybrid approach has the further advantages that it puts the echo into a useful context by positioning it in a familiar coordinate system and the background X-ray helps in identifying catheters and devices in the lower quality echo image. Future work will focus on automatic detection of registration failures in order to maximise clinical robustness.

References

1. Cleary, K., Peters, T.M.: Image-guided interventions: technology review and clinical applications. Annual Review of Biomedical Engineering 12, 119–142 (2010)
2. Jain, A., Gutierrez, L., Stanton, D.: 3D TEE Registration with X-Ray Fluoroscopy for Interventional Cardiac Applications. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 321–329. Springer, Heidelberg (2009)
3. Ma, Y., Penney, G.P., Bos, D., Frissen, P., Rinaldi, C.A., Razavi, R., Rhode, K.S.: Hybrid echo and X-ray image guidance for cardiac catheterization procedures by using a robotic arm: a feasibility study. Physics in Medicine and Biology 55(13), N371–N382 (2010)
4. Gao, G., Penney, G., Ma, Y., Gogin, N., Cathier, P., Arujuna, A., Morton, G., Caulfield, D., Gill, J., Rinaldi, C.A., Hancock, J., Redwood, S., Thomas, M., Razavi, R., Gijsbers, G., Rhode, K.: Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking. Medical Image Analysis 16(1), 38–49 (2012)
5. Lang, P., Seslija, P., Habets, D.F., Chu, M.W.A., Holdsworth, D.W., Peters, T.M.: Three-Dimensional Ultrasound Probe Pose Estimation from Single-Perspective X-Rays for Image-Guided Interventions. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 344–352. Springer, Heidelberg (2010)
6. van de Kraats, E.B., Penney, G.P., Tomazevic, D., van Walsum, T., Niessen, W.J.: Standardized evaluation methodology for 2-D-3-D registration. IEEE Transactions on Medical Imaging 24(9), 1177–1189 (2005)
7. Linte, C.A., Lang, P., Rettmann, M.E., Cho, D.S., Holmes, D.R., Robb, R.A., Peters, T.M.: Accuracy considerations in image-guided cardiac interventions: experience and lessons learned. International Journal of Computer Assisted Radiology and Surgery 7(1), 13–25 (2012)

LBM-EP: Lattice-Boltzmann Method
for Fast Cardiac Electrophysiology Simulation
from 3D Images

S. Rapaka[1], T. Mansi[1], B. Georgescu[1], M. Pop[2], G.A. Wright[2],
A. Kamen[1], and Dorin Comaniciu[1]

[1] Siemens Corporation, Corporate Research and Technology,
Imaging and Computer Vision, Princeton, NJ, USA
[2] Department of Medical Biophysics, University of Toronto,
Sunnybrook Health Sciences Centre, Imaging Research, Toronto, ON, Canada

Abstract. Current treatments of heart rhythm troubles require careful planning and guidance for optimal outcomes. Computational models of cardiac electrophysiology are being proposed for therapy planning but current approaches are either too simplified or too computationally intensive for patient-specific simulations in clinical practice. This paper presents a novel approach, LBM-EP, to solve any type of mono-domain cardiac electrophysiology models at near real-time that is especially tailored for patient-specific simulations. The domain is discretized on a Cartesian grid with a level-set representation of patient's heart geometry, previously estimated from images automatically. The cell model is calculated node-wise, while the transmembrane potential is diffused using Lattice-Boltzmann method within the domain defined by the level-set. Experiments on synthetic cases, on a data set from CESC'10 and on one patient with myocardium scar showed that LBM-EP provides results comparable to an FEM implementation, while being $10 - 45$ times faster. Fast, accurate, scalable and requiring no specific meshing, LBM-EP paves the way to efficient and detailed models of cardiac electrophysiology for therapy planning.

1 Introduction

Since the seminal work of Hodgkin and Huxley [5], a large variety of models have been proposed to simulate the propagation of the action potential (AP) across the heart muscle, with various degrees of complexity ([3]). Biophysical models aim to capture the ion interactions and protein mechanisms that regulate the AP. At a higher scale, phenomenological models have been developed to mimic the AP without directly considering the underlying molecular mechanisms. Finally, Eikonal models do not simulate the AP altogether but the propagation of the electrical front directly. Since these models are generative, they may constitute efficient tools for therapy planning. Phenomenological models provide a good compromise between model details, being able to capture most of the pathological conditions, and complexity. Recent works demonstrated that those models

can be personalised from clinical data [1]. However, because they solve stiff partial differential equations (PDE), they are still too computationally demanding for day-to-day clinical setups and intervention guidance. Another limitation is the requirement of high-quality/high-resolution volume meshes, which can be difficult to obtain from patient-specific anatomies.

In recent years, the lattice-Boltzmann method (LBM) [2] has developed as a powerful technique for accurate simulation of a large class of partial-differential equations. In particular, it has been successfully applied to pattern-forming reaction-diffusion equations([4]). While originally developed from cellular automata models of fluid flows, the method has found a firm theoretical basis in kinetic theory of weakly-compressible fluid flows. Some of the key strengths of this method are, *i)* local nature of the computational algorithm, which provides very high scalability on modern parallel computing architectures, *ii)* second-order accuracy in space, and *iii)* simplicity of implementation on a uniform Cartesian grid.

Motivated by the recent breakthrough in LBM, we present a novel framework, henceforth called LBM-EP, for efficient patient-specific simulations of cardiac electrophysiology models at near real-time. Although general, the method is illustrated in this study on the Mitchell-Schaeffer model [6]. Sec. 2 presents a description of the algorithms used to process the medical images, and the lattice-Boltzmann algorithm used for propagating the action-potential. Sec. 3 compares the simulation results computed with the proposed LBM-EP with an FEM implementation of Mitchell-Schaeffer model in synthetic scenarios, a dataset of CESC'10 Grand Challenge and one patient with a myocardium scar, showing an accuracy in the range of the variability reported in the literature and a speed-up of about $10 - 45\times$ with respect to FEM. Sec. 4 concludes the paper.

2   Methods

2.1   Computational Domain Preparation from Medical Images

LBM-EP being solved on Cartesian grids (Sec. 2.3), its application to clinical images is relatively immediate. Starting from a cardiac image (e.g. cine MRI), the left endocardium, right endocardium and epicardium are automatically segmented using a machine learning approach [11] and fused in one surface representing the myocardium while preserving their anatomical label. A level-set representation of that surface is then computed on an isotropic Cartesian grid. Based on the labels, grid nodes lying at the heart endocardia are marked for electrophysiology stimulation and synthetic fibers are computed by linearly interpolating the elevation angle from $-70°$ at the epicardium to $+70°$ at the endocardium [3]. Scars can be reported in the domain through level set (Sec. 3.3).

2.2   Mitchell-Schaeffer Model of Action Potential

Mitchell-Schaeffer (M-S) model [6] is employed here although the method can be extended to other mono-domain models. The model (Eq. (1)) relates the

normalized transmembrane AP $v(t) \in [0,1]$ to an inward gated current $J_{in} = hv^2(1-v)/\tau_{in}$, which captures the fast acting currents using the gating variable $h(t)$, and an outward un-gated current $J_{out} = -v/\tau_{out}$, which accounts for transmembrane voltage decrease. A transient stimulus current J_{stim} is added to the model to simulate electrical pacing.

$$\frac{\partial v}{\partial t} = J_{int} + J_{out} + J_{stim} + c\nabla \cdot D\nabla v \tag{1}$$

The gating variable $h(t)$ evolves according to $\frac{dh}{dt} = \frac{1-h}{\tau_{open}}$ if $v < v_{gate}$, otherwise $\frac{dh}{dt} = \frac{-h}{\tau_{close}}$ where v_{gate} is the change-over voltage, $D = \rho\text{Id} + (1-\rho)\mathbf{aa}^\intercal$ is the anisotropic diffusion tensor along the fiber direction $\mathbf{a}$ with anisotropy ratio ρ, c is the diffusion coefficient along the fibers. τ_{in}, τ_{out}, τ_{open} and τ_{close} are directly related to the shape and duration of the AP ([6]), which makes their calibration possible from clinical data.

2.3 Lattice-Boltzmann Model of Cardiac Electrophysiology

Eq. (1) is solved on the Cartesian grid (Sec. 2.1) using Lattice-Boltzmann method with a 7-connectivity topology (6 connections + central position) and Neumann boundary conditions. The gating variable $h(t)$ is updated at every node of the grid using a forward Euler scheme. The fundamental variable of LBM is the vector of distribution functions $\mathbf{f}(\mathbf{x}) = \{f_i(\mathbf{x})\}_{i=1\ldots7}$, where $f_i(\mathbf{x})$ represents the probability of finding a particle travelling along the edge $\mathbf{e}_i$ of node $\mathbf{x}$. The governing equation at $\mathbf{x}$ for the edge $\mathbf{e}_i$ is composed of two successive steps:

$$f_i^* = f_i - A_{ij}(f_j - \omega_j v) + \delta t\omega_i(J_{in} + J_{out} + J_{stim}), \tag{2}$$
$$f_i(\mathbf{x} + \mathbf{e}_i, t + \delta t) = f_i^*(\mathbf{x}, t) \tag{3}$$

where, the collision matrix $A = (A_{ij})_{i,j\in[\![1,7]\!]}$ relaxes the distribution function f_i towards the local value of the potential, v, f_i^* is an intermediate, post-collision state of the distribution function, and ω_i is a weighting factor that depends on lattice connectivity, here $\omega_i = 1/8$ for the edges to the six neighbors and $\omega_i = 1/4$ for the central position. The transmembrane AP is related to the f_i's through $v(\mathbf{x}, t) = \sum_i f_i(\mathbf{x}, t)$. For each time step δt, a strictly local collision rule (Eq. (2)) is applied to the distribution functions at each node. Post-collision, the distribution functions stream along their corresponding edges to the neighboring nodes (Eq. (3)). In its simplest form, the collision matrix is diagonal with a characteristic relaxation time τ, $A = (1/\tau)I$, where I is the 7×7 identity matrix.

At the problem boundaries, the streaming step requires the specification of additional incoming distribution functions to ensure proper boundary conditions. It can be shown that the potential gradient at a node is related to the f_i's through $c\nabla v = (1-1/2\tau)\sum_i f_i\mathbf{e}_i$ [4]. The Neumann boundary condition for potential on a surface simplifies to $\sum_i f_i\mathbf{e}_i \cdot \mathbf{n} = 0$. If the boundary is normal to any edge of the lattice, the Neumann boundary condition is automatically recovered if the incoming distribution at the node is equal to the outgoing one. Complex geometries can be handled easily by means of a level set formulation. The incoming

distribution is calculated automatically from the distance to the wall at nodes close to the boundary as provided by the level-set [10], thus enabling simulations in complex domain without requiring advanced meshing algorithms.

Remarkably, this simple model can be shown to reproduce the reaction-diffusion equation Eq. (1) with an isotropic diffusion coefficient of $c = (2\tau - 1)/8$ (see [4] for the proof). To extend the model for anisotropic diffusion, like in the heart, the matrix A is replaced by $A = M^{-1}SM$ [9], where

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & -6 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -2 & -2 & 0 \end{pmatrix} \quad S^{-1} = \begin{pmatrix} \tau_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \tau_{11} & \tau_{12} & \tau_{13} & 0 & 0 & 0 \\ 0 & \tau_{21} & \tau_{22} & \tau_{23} & 0 & 0 & 0 \\ 0 & \tau_{31} & \tau_{32} & \tau_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \tau_5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \tau_6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \tau_7 \end{pmatrix}$$

The first row of M corresponds to $v = \sum_i f_i$, while rows $2 - 4$ are the three components of the potential gradient. The relaxation times $(\tau_{ij})_{i,j \in [\![1,3]\!]}$ are related to the components of the diffusion tensor through $\tau_{ij} = \delta_{ij}/2 + 4D_{ij}\delta t/\delta x^2$. The relaxation times τ_1 related to the potential and τ_5, τ_6 and τ_7 related to the higher order moments do not directly effect the diffusion solution, but effect the stability of the method. In this work, we use $\tau_1 = 1$ and $\tau_5 = \tau_6 = \tau_7 = 1.33$. Algo. 1 summarizes the main steps of the method.

Algorithm 1. LBM-EP: Lattice-Boltzmann Model of Cardiac Electrophysiology

Require: Cartesian grid, level-set domain boundaries, δt, $nbIter$, model parameters
1: **for** $iter = 1 \to nbIter$ **do**
2:     $t \leftarrow t + \delta t$
3:     **for** every node **x do**
4:         $\forall i$, compute post-collision distributions $f_i^*(\mathbf{x})$ (Eq. 2)
5:         Update $h(\mathbf{x})$
6:     **for** every node **x do**
7:         $\forall i$, stream $f_i(\mathbf{x})$ and apply boundary conditions (Eq. 3)
8: **return** $v = \sum_i f_i$, h.

---

3   Experiments and Results

All experiments were executed on a standard Windows XP desktop machine (Intel Xeon, 2.40GHz octo-core, 4GB RAM). LBM-EP was implemented in Fortran with no particular optimization. A semi-implicit, anisotropic finite element implementation of M-S model, called FEM-EP, was used for comparisons. FEM-EP was based on linear tetrahedra and parallel optimization (OpenMP).

3.1   Quantitative Evaluation on Synthetic Scenarios

We first compared the performance of LBM-EP with respect to FEM-EP in two different scenarios representing the main pathological features. To that end, a $10 \times 10 \times 0.5 \, cm$ slab was discretized into $401 \times 401 \times 3$ nodes (1,920,000 tetrahedra for FEM-EP). For both LBM-EP and FEM-EP, we used $\delta t = 0.1 ms$, $\tau_{close} = 150 \, ms$, $\tau_{open} = 120 \, ms$, $\tau_{in} = 0.3 \, ms$, $\tau_{out} = 6 \, ms$, $v_{gate} = 0.13$,

$c = 0.0003\,cm^2/ms$ and $\rho = 1$ (isotropic diffusion) [6]. An electrical stimulation was applied at the nodes $(x, y = 0, z)$ for $1\,ms$ duration by setting $v = 1$ at these nodes. Grid resolution analysis (not reported here) showed that numerical convergence was reached for both models at that spatial and temporal resolution.

We first tested the ability of LBM-EP to capture front-bending around a scar. A scar region was simulated within the domain (Fig. 1, left panels) by: setting the diffusion coefficient $c = 0$ for FEM-EP, defining Neumann boundary conditions for LBM-EP. From the computed depolarization times (Fig. 1, left panels), one can see that both models yielded very similar behavior. The AP front calculated by LBM-EP correctly rotated around the scar. Where the front was not perturbed by the scar, both models yielded nearly identical results, as quantified by the AP at Point 2 ($0.3\,ms$ difference in depolarization time, Fig. 1, right panel). Near the scar, slight differences could be identified, which resulted in a difference of $9.7\,ms$ in depolarization time at Point 1. However, this cannot be interpreted as a limitation of LBM-EP, which captured perfectly the Neumann boundary conditions around the scar, contrary to FEM-EP.
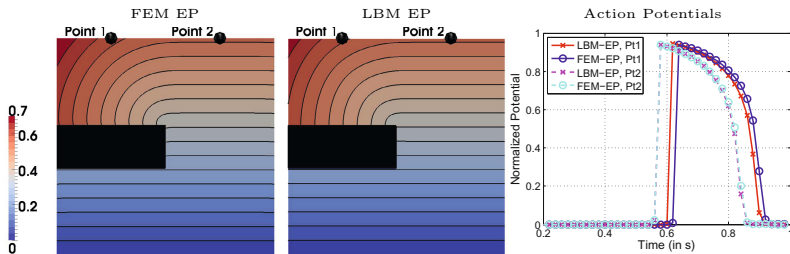
Fig. 1. *Left panels*: Computed depolarization times (in ms) in a homogeneous medium with scar (*in black*). LBM-EP captured front rotation around the scar. *Right panel*: action potential at points 1 and 2, showing consistent results between both models.

We then tested the ability of LBM-EP to simulate vortex formation due to premature stimulation. For that, we set the diffusion coefficient to $0.0012\,cm^2/ms$, removed the scar and applied a second stimulation at the nodes ($x \leq 0.5, y = 0.5, z$) at $t = 452\,ms$ for both methods. As shown in Fig. 2, the patterns obtained with LBM-EP were similar to those obtained with FEM-EP. LBM-EP could be used to simulate complex pathologies like fibrillation or tachycardia.

Computation Time. For all experiments, FEM-EP required $\approx 700\,ms$ per iteration whereas LBM-EP required only $\approx 80\,ms$, about $8.75\times$ speed-up.

3.2   Comparison with Published Results on CESC'10 Data

We evaluated LBM-EP performance in a dataset distributed during CESC'10 MICCAI Grand Challenge with respect to FEM-EP and to a recently published benchmark [1]. Our purpose being evaluation and not personalization, we did not adjust the parameters locally. We thus compared our results with the generic benchmark of the ionic ten Tusscher-Panfilov model only [1]. CESC'10 dataset
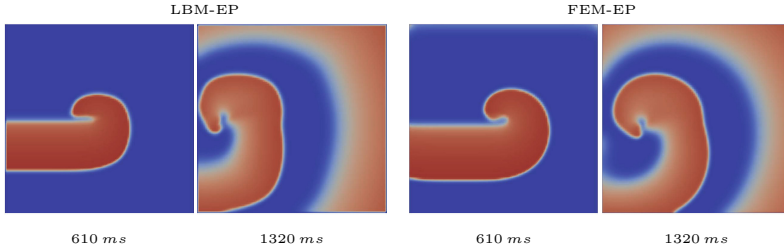
Fig. 2. LBM-EP succesfully captured vortex formation due to premature stimulation

consisted in an explanted porcine heart, and comprised optical fluorescence images of transmembrane potential and high-resolution diffusion-weighted (DW) MRI images ([7]). The optical images contained the depolarization and repolarization phase of transmembrane potential under left endocardium and right epicardium pacings at $1.1\,Hz$. Finally, the mesh constructed from DW-MRI had the fiber directions integrated, providing a complete model of the heart anatomy.

LBM-EP was computed on $0.5\,mm$ grid while FEM-EP was computed on the provided tetrahedra mesh with $0.5\,mm$ average edge-length. The time steps were set to $\delta t_{LBM-EP} = 0.1\,ms$ and $\delta t_{FEM-EP} = 0.5\,ms$. Myocardium fibers defined on the tetrahedra mesh were rasterized on the LBM lattice for simulation. The generic M-S parameters were used for both models [6], with a diffusion coefficient $c = 0.0035\,cm^2/ms$ and anisotropy ratio $\rho = 0.25$. As it can be seen from Fig. 3, LBM-EP simulation was qualitatively similar to FEM-EP for both pacing conditions in terms of depolarization isochrone patterns. The difference in absolute depolarization time was mostly due to the different computational domain, in particular regarding the precise location of the excitation nodes and the fiber orientation, which was locally altered by the rasterization. Compared to the CESC'10 benchmark, LBM-EP provided similar depolarization patterns, suggesting promising reproducibility and validity. Finally, while FEM-EP required $\approx 16\,s$ per iteration, LBM-EP took only $\approx 0.35\,s$ per iteration, which corresponds to a speed-up of $45\times$. It should be noted that the computational efficiency of our FEM implementation was similar to those reported in the literature, $\approx 1\,s$/iteration on the $1.5\,mm$ mesh provided by the challengers as in [8].

3.3   Real Case Example

We finally illustrate how LBM-EP can be used in a real clinical scenario on a patient with myocardium scar due to previous surgery. Parameters were kept generic as no electrophysiology data were available. Fast conductivity was modeled on the endocardium to mimic the Purkinje fibers. The scar was represented as a level-set to ensure Neumann boundary conditions. Fig. 4 shows the depolarization time isochrones, illustrating some delays at the apex due to the scar. Elsewhere in the myocardium the isochrones presented with patterns similar to what has been reported in the literature [3]. For this patient, one time step was calculated in $0.2\,s$ for a grid size of $1\,mm$.
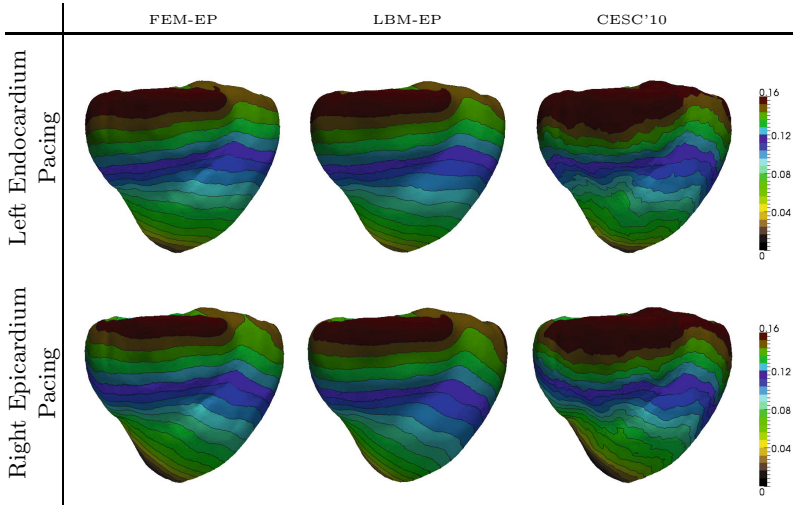
Fig. 3. FEM-EP and LBM-EP simulations using Mitchell-Schaeffer model and CESC'10 data using ten Tusscher-Panfilov model (see text for details)
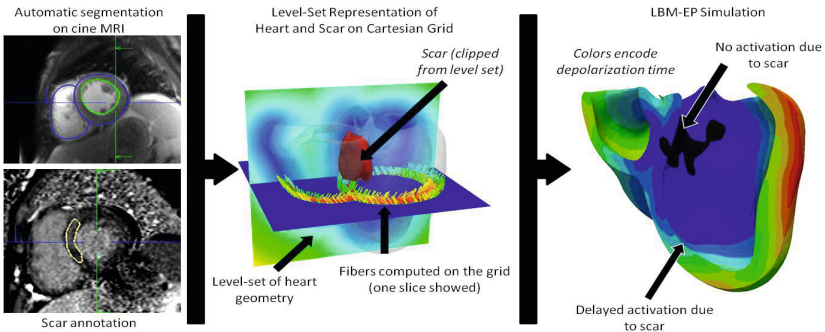
Fig. 4. LBM-EP simulation on one patient with myocardium scar

4   Conclusion and Future Work

We have presented in this paper, to the best of our knowledge, the first application of a near-real time lattice-Boltzmann model for general monodomain models of cardiac electrophysiology. Node-based by construction, our framework does not require advanced meshing and can be applied directly from images by means of level-sets. Through comparisons with the traditional finite-element method, we have empirically shown the applicability of LBM-EP to cardiac electrophysiology. A comprehensive description of the theory along with detailed studies of convergence and accuracy are being prepared as a larger contribution. LBM-EP provides second order accuracy in space, can be easily extended to any type of mono-domain cellular model, and, above all, offers between $10 - 45\times$ speed-up with respect to traditional FEM. Preliminary experiments on graphi-

cal processing units (GPUs) showed the potential for another order-of-magnitude improvement in computational speed. This is the first time, to the best of our knowledge, that very fast simulations of cardiac electrophysiology is achieved with a detailed model. Our method may thus constitute an ideal framework for patient-specific simulations for therapy planning and real-time guidance.

References

1. Camara, O., Sermesant, M., Lamata, P., Wang, L., Pop, M., Relan, J., De Craene, M., Delingette, H., Liu, H., Niederer, S., Pashaei, A., Plank, G., Romero, D., Sebastian, R., Wong, K.C., Zhang, H., Ayache, N., Frangi, A., Shi, P., Smith, N., Wright, G.: Inter-Model Consistency and Complementarity: Learning from ex-vivo Imaging and Electrophysiological Data towards an Integrated Understanding of Cardiac Physiology. Prog. Biophys. Mol. Biol. 107, 122–133 (2011)
2. Chen, S., Doolen, G.D.: Lattice Boltzmann method for fluid flows. Annu. Rev. Fluid Mech. 30(1), 329–364 (1998)
3. Clayton, R.H., Bernus, O., Cherry, E.M., Dierckx, H., Fenton, F.H., Mirabella, L., Panfilov, A.V., Sachse, F.B., Seemann, G., Zhang, H.: Models of cardiac tissue electrophysiology: Progress, challenges and open questions. Prog. Biophys. Mol. Biol. 104(1-3), 22 (2011)
4. Dawson, S.P., Chen, S., Doolen, G.D.: Lattice Boltzmann computations for reaction-diffusion equations. J. Chem. Phys. 98(2), 1514–1523 (1993)
5. Hodgkin, A.L., Huxley, A.F.: A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. J. Physiol. 117(4), 500–544 (1952)
6. Mitchell, C.C., Schaeffer, D.G.: A two-current model for the dynamics of cardiac membrane. Bull. Math. Biol. 65(5), 767–793 (2003)
7. Pop, M., Sermesant, M., Lepiller, D., Truong, M.V., McVeigh, E.R., Crystal, E., Dick, A., Delingette, H., Ayache, N., Wright, G.A.: Fusion of optical imaging and mri for the evaluation and adjustment of macroscopic models of cardiac electrophysiology: A feasibility study. Med. Image Anal. 13(2), 370–380 (2009)
8. Relan, J., Sermesant, M., Pop, M., Delingette, H., Sorine, M., Wright, G., Ayache, N.: Volumetric Prediction of Cardiac Electrophysiology using a Heart Model Personalised to Surface Data. In: CI2BM 2009 - MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling (2009)
9. Yoshida, H., Nagaoka, M.: Multiple-relaxation-time lattice Boltzmann model for the convection and anisotropic diffusion equation. J. Comp. Phys. 229, 7774–7795 (2010)
10. Yu, D., Mei, R., Luo, L., Shyy, W.: Viscous flow computations with the method of lattice Boltzmann equation. Prog. Aero. Sci. 39(5), 329–367 (2003)
11. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. Med. Imaging 27(11), 1668–1681 (2008)

Cardiac Mechanical Parameter Calibration Based on the Unscented Transform

Stéphanie Marchesseau[1,*], Hervé Delingette[1], Maxime Sermesant[1],
Kawal Rhode[2], Simon G. Duckett[2], C. Aldo Rinaldi[2],
Reza Razavi[2], and Nicholas Ayache[1]

[1] Asclepios Research Project, INRIA Sophia Antipolis, France
[2] King's College London, Division of Imaging Sciences & Biomedical Engineering,
St. Thomas' Hospital, London, UK

Abstract. Patient-specific cardiac modelling can help in understanding pathophysiology and predict therapy planning. However it requires to personalize the model geometry, kinematics, electrophysiology and mechanics. Calibration aims at providing global values (space invariant) of parameters before performing the personalization stage which involves solving an inverse problem to find regional values. We propose an automatic calibration method of the mechanical parameters of the Bestel-Clément-Sorine (BCS) electromechanical model of the heart based on the Unscented Transform algorithm. A sensitivity analysis is performed that reveals which observations on the volume and pressure evolution are significant to characterize the global behaviour of the myocardium. We show that the calibration method gives satisfying results by optimizing up to 7 parameters of the BCS model in only one iteration. This method was evaluated on 7 volunteers and 2 heart failure patients, with a mean relative error from the real data of 11%. This calibration enabled furthermore a preliminary study of the specific parameters to the studied pathologies.

1 Introduction

Patient-specific cardiac modelling can provide additional guidance to cardiologists in understanding pathophysiology and predict therapy planning. Several approaches for the past 20 years have been developed to describe and simulate the cardiac function, including cardiac mechanics and electrophysiology [4,1]. They differ in their choice of hyperelastic material, electrophysiological properties or electromechanical coupling. In this paper the Bestel-Clement-Sorine (BCS) model [1], further improved by [3] is used.

The simulation becomes patient-specific after several levels of personalization: geometrical, kinematic, electrophysiological and mechanical. Mechanical personalization consists in optimizing mechanical parameters of the model so that the

---

simulation behaves in accordance to patient-specific datasets (images and other signals).

This inverse problem has been tackled by different authors (for instance [12], [7],[8] or [10]). However there is no guarantee that such algorithms will converge toward a relevant solution due to their dependence on an initial range of parameter values. The choice of the parameters to estimate and their initial calibration has therefore great impact for the personalization.

Our main contribution tackles this initialization issue: we propose a simple and efficient method to automatically calibrate the parameters from the ventricular volume or pressure evolution over the cardiac cycle. It has been applied successfully for the calibration of mechanical parameters from 7 healthy cases and has been tested in two heart failure cases. Our proposed method is based on the Unscented Transform algorithm and requires only one iteration with multiple simulations performed in parallel for calibrating up to 7 parameters selected from a sensitivity analysis. Moreover, a comparison between the estimated parameters for control and heart failure cases enabled a preliminary specificity study that aims at classifying the pathologies.

2   The Bestel-Clement-Sorine Electromechanical Model of the Heart

Our approach is based on the Bestel-Clement-Sorine (BCS) model [1] further improved by [3]. The model is composed of a passive isotropic visco-hyperelastic component that accounts for the elasticity and the friction in the cardiac extracellular matrix surrounding the fibres, described as a Mooney-Rivlin material. The strain energy for a Mooney-Rivlin material is given as: $W_e = c_1(\bar{I}_1 - 3) + c_2(\bar{I}_2 - 3) + \frac{K}{2}(J - 1)^2$, where c_1, c_2 are material parameters and K is the Bulk modulus. The quantities $\bar{I}_1$ and $\bar{I}_2$ are the isochoric invariants of the *Cauchy-deformation tensor* $\mathbf{C}$.

In parallel, the stress along the cardiac fibre is composed of an active part (contraction in the sarcomere) and a passive part corresponding to the elastic bound (titin) between sarcomeres and Z-discs, having stress $\sigma_s = E_s e_s$. The contractile component having stress tensor σ_c, driven by the control variable u, has a viscous part to account for the energy dissipated in the sarcomere due to friction. This gives $\sigma_c = \tau_c + \mu \dot{e}_c$. Fig. 1 shows a rheological representation of this model.

At the nanoscopic scale, the binding and unbinding process of the actin and myosin filaments in the sarcomere is described by Huxley's filament model [5]. Statistical mechanics allows to describe its behavior at the macroscopic scale, resulting in a differential equation that controls the active stress τ_c and the sarcomere stiffness k_c:

$$\begin{cases} \dot{k_c} = -(\mid u \mid +\alpha \mid \dot{e}_c \mid)k_c + n_0 k_0 \mid u \mid_+ \\ \dot{\tau_c} = -(\mid u \mid +\alpha \mid \dot{e}_c \mid)\tau_c + \dot{e}_c k_c + n_0 \sigma_0 \mid u \mid_+ \end{cases} \quad (1)$$

where α is a constant related to the cross-bridge release due to a high contraction rate, k_0 and σ_0 are respectively the maximum stiffness and contraction. n_0 is
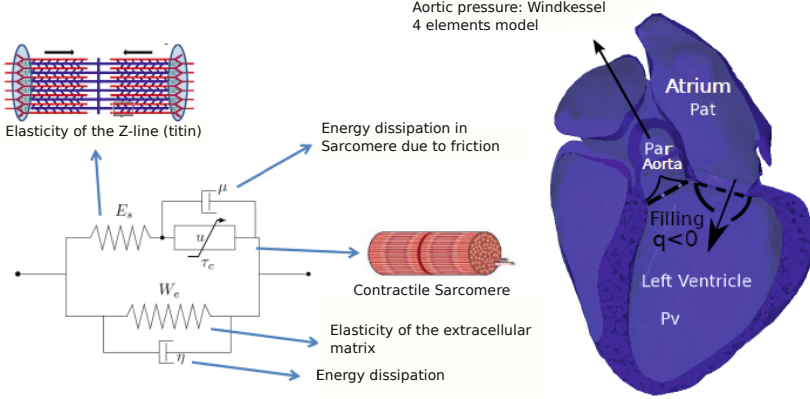
Fig. 1. Full electromechanical and circulation model. (Left) W_e is the strain energy of the extracellular matrix considered here as an isotropic material, associated with a dissipative term η. u is a control variable which is driven by changes in transmembrane potential. It controls the contraction stress τ_c. μ deals with the friction in the sarcomere while E_s is a linear spring to enforce elasticity of the Z-discs (titin). (Right) Circulation model in the filling phase for the left ventricle.

a reduction factor that allows to take into account *the Starling effect* by which the maximum contraction depends on the fibre strain e_c. The control variable u is derived from the electrical activation model and is a function of the free calcium concentration only. It is modeled using electrophysiological inputs such as depolarization times (T_d) and action potential durations (APD) and depends on two parameters: k_{ATP} the rate of the myosin ATPase activity controlling the contraction rate and k_{RS} the rate of sarcoplasmic reticulum calcium re-uptake controlling the relaxation rate.

The ventricles are filled with blood coming from the atria and ejected through the arteries. A basic circulation model is represented in Fig. 1. A valve model explained in [3] gives relationships between the outward flow and the various pressures (ventricular, arteria and atria). The arteria pressure is modeled using a four-element Windkessel model [9], that depends on four parameters: the peripheral resistance R_p, the characteristic time τ, the characteristic resistance Z_c and the total arteria inertance L.

3   Unscented Transform-Based Parameter Calibration

To calibrate the model, we use the algorithm derived from the Unscented Transform [6]. We chose the ventricular volume curves as main observations to perform the calibration as they are important physiological indices and can be captured by few quantities: the minimum volume V_{min}, the maximum and minimum of the flow (q_{max} and q_{min} respectively). Moreover, if available, the maximum pressure

P_{max}, as well as the maximum and minimum of the derivative of the pressure (dP/dt_{max} and dP/dt_{min} respectively) can be taken into account.

3.1   Unscented Transform Algorithm

The proposed algorithm finds a set of parameters that minimize the difference between the measured observation $\mathbf{Z}^{obs}$ and the predicted observation $\bar{\mathbf{Z}}$. It is explained as follows: Let $\mathbf{Z}$ be the vector of observations and $\mathbf{X}$ the parameter vector which has mean $\mathbf{X}^0$, covariance $\mathbf{C}_X$ and dimension n. We set the covariance as $\mathbf{C}_X = \mathbf{Cov}(\mathbf{X}, \mathbf{X})$ by estimating the minimal and the maximal value of each parameter with a trial and error approach on one volunteer. We compute observations $\mathbf{Z}_{i\epsilon}$ from the $2n + 1$ sets of parameters $\mathbf{X}_{i\epsilon} = [x_1, x_2, ..., x_i + \epsilon s_i, ...]$ around the mean value $\mathbf{X}^0$ where $\epsilon \in \{-1, 0, 1\}$ and s_i is an uncertainty function of the covariance $s_i = \gamma\sqrt{\mathbf{C}_{X_i}}$, with γ the scaling parameters. The mean observation is set as $\bar{\mathbf{Z}} = \sum_{i,\epsilon} \omega_{i\epsilon} \mathbf{Z}_{i\epsilon}$ with some weights $\omega_{i\epsilon}$ described by [11]. Finally we derive the covariance matrix as:

$$\mathbf{Cov}(\mathbf{X}, \mathbf{Z}) = \sum_{i\epsilon} \omega_{i\epsilon}(\mathbf{X}_{i\epsilon} - \mathbf{X}^0)\ (\mathbf{Z}_{i\epsilon} - \bar{\mathbf{Z}})^T \tag{2}$$

The new set of parameters $\mathbf{X}^{new}$ found to match the observations $\mathbf{Z}^{obs}$ is

$$(\mathbf{X}^{new} - \mathbf{X}^0) = \mathbf{Cov}(\mathbf{X}, \mathbf{Z})\ \mathbf{Cov}(\mathbf{Z}, \mathbf{Z})^{-1}\ (\mathbf{Z}^{obs} - \bar{\mathbf{Z}}) \tag{3}$$

where

$$\mathbf{Cov}(\mathbf{Z}, \mathbf{Z}) = \sum_{i\epsilon} \omega_{i\epsilon}(\mathbf{Z}_{i\epsilon} - \bar{\mathbf{Z}})\ (\mathbf{Z}_{i\epsilon} - \bar{\mathbf{Z}})^T. \tag{4}$$

This algorithm is very simple to implement and runs in one iteration to give $\mathbf{X}^{new}$. Another simulation is necessary to obtain the resulting observation $\mathbf{Z}^{new}$.

3.2   Parameter Selection

Fourteen parameters in total have to be estimated: $(\sigma_0, k_{rs}, k_{atp}, k_0, \alpha, \mu, E_s)$ active parameters, (K, c_1, c_2) passive parameters and (R_p, τ, Z_c, L) for the valve model. Since it is not reasonable to try to estimate all of them at once, we decide to fix some to a standard value and estimate others. A sensitivity analysis was performed in order to select the main parameters. We chose the following four parameters (σ_0, μ, K, R_p) that we might be able to estimate from a volume curve. This choice was confirmed by a Singular-Value-Decomposition (SVD) of the covariance matrix $\mathbf{Cov}(\mathbf{X}, \mathbf{Z})$ made from all fourteen parameters and the three observations (V_{min}, q_{max} and q_{min}). When pressure curves are available, more parameters can be estimated. k_{rs}, k_{atp} as well as the stiffness parameter c_1 were chosen since they greatly influence the pressure slopes. Relevant curves are presented in Fig. 2.
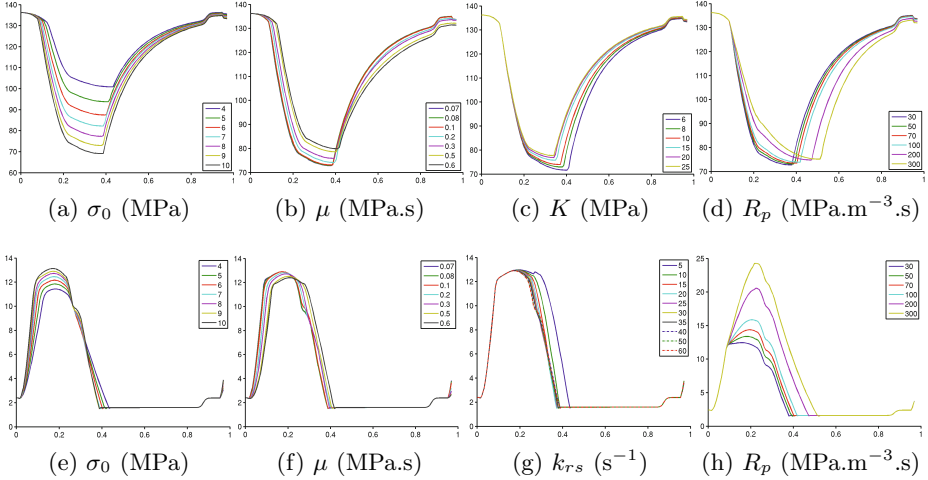
Fig. 2. Observations over time: (Top) Volume in mL, (Bottom) Pressure in kPa

4   Calibration Results on Healthy and Pathological Cases

We demonstrate the application of the proposed method on cardiac MRI data
on both volunteer and patient data acquired at the Division of Imaging Sciences
& Biomedical Engineering at King's College London, UK. Initial parameters X^0
and covariance matrix $\mathbf{C}_X$ are the same for all cases. Each case was calibrated in
about 20 minutes that includes the time to run in parallel the $2n+1$ simulations,
the calibration time and the final simulation with the calibrated parameters.

4.1   Volunteer Data: Calibration with Volume Curves

The study was performed on seven healthy hearts provided by the STACOM
challenge. The electrophysiological model was simulated with standard values
and healthy onset (see Fig.3 Left). From the kinematic personalization, we reg-
istered all images on the end diastolic image. Then, image transformations were
applied to the end-diastolic tetrahedral mesh to estimate the volume of the
ventricles over time and then the observation vector $\mathbf{Z}^{obs}$. Fig. 3 shows the mea-
sured, reference and estimated volume curves on case 3 and errors between the
real observations $\mathbf{Z}^{obs}$ and the simulated observations $\mathbf{Z}^{new}$ are given in Table 1.
Moreover, the calibration provided a consistent and plausible range of global
values for parameters that will be used in a specificity study (see Table 2).

4.2   Pathological Data: Calibration with Volume and Pressures

The proposed calibration approach was applied on two Left Bundle Branch Block
(LBBB)[1] cases. Electrophysiological personalization (see Fig. 4 left) was enabled

---

[1] LBBB cases are characterized by dyssynchronous electrophysiology.

Table 1. Relative errors (in %) between simulated results $\mathbf{Z}^{new}$ and real data $\mathbf{Z}^{obs}$ on the 7 healthy cases

Volunteers	1	2	3	4	5	6	7	Mean
V_{min}	0.35	3.51	0.83	0.79	1.09	1.38	1.31	1.32
q_{max}	3.06	20.99	8.57	21.37	11.5	12.1	5.36	11.85
q_{min}	0.31	4.12	27.41	6.48	27.58	16.92	5.81	12.66

Table 2. Estimated parameters from the calibration

Volunteers	1	2	3	4	5	6	7	Mean
σ_0 (MPa)	6.49	4.42	4.92	5.46	5.51	8.75	5.32	5.8
μ (MPa.s)	0.31	0.31	0.27	0.3	0.33	0.26	0.3	0.3
K (MPa)	14.22	10.44	12.72	13.24	14.12	12.72	12.82	12.9
R_p (MPa.m^{-3}.s)	93.87	130.88	110.1	116.73	104.43	98.3	141.39	113.7
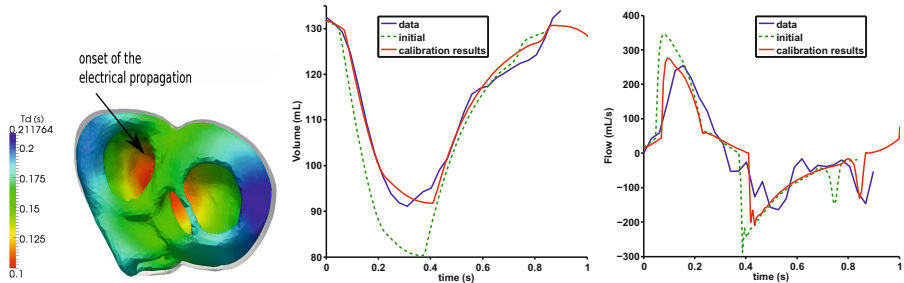
Fig. 3. (Left) Electrophysiological and geometrical input. (Right) Results of the calibration technique on real data for one healthy volunteer.

and pressure curves were available thanks to pressure sensors. Results on the two LBBB cases are given in Fig. 4 and the comparison with the images is given in Fig. 5. The mean relative error for both cases is about 16%. We noticed a much higher passive stiffness ($K = 19$MPa) for LBBB case 2 which also suffers from dilated cardiomyopathy, and a smaller contractility ($\sigma_0 < 4$MPa) for both heart failure cases. These results are in agreement with medical knowledge (see [2] for instance). Therefore we can assess that a small contractility may be specific to LBBB cases as a high myocardial stiffness may be to dilated cardiomyopathy.

4.3 Evaluation of Registration Error Influence

We tried to evaluate the error in the registration technique to understand whether the model could actually match the data better than shown in Fig. 3. For this purpose, we created synthetic images from a real sequence, using the deformed meshes resulting from a simulation. We then registered this new sequence with the same registration technique as used for the kinematics personalization, and extracted the volume curves from the resulting registered meshes. The comparison between
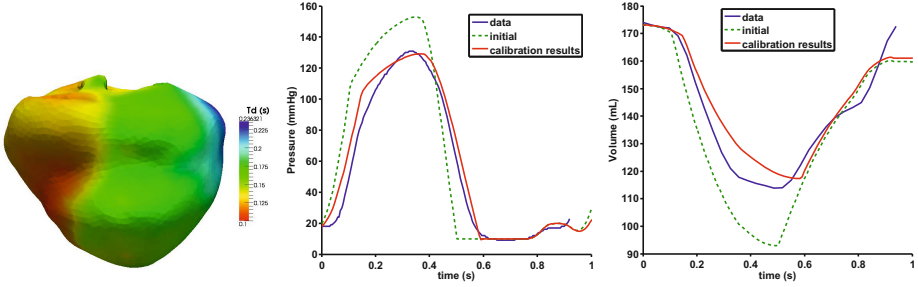
Fig. 4. (Left) Personalized electrophysiology. (Right) Results of the calibration technique on real data for LBBB case 1.

(a) t = 0 ms

(b) t = 410 ms
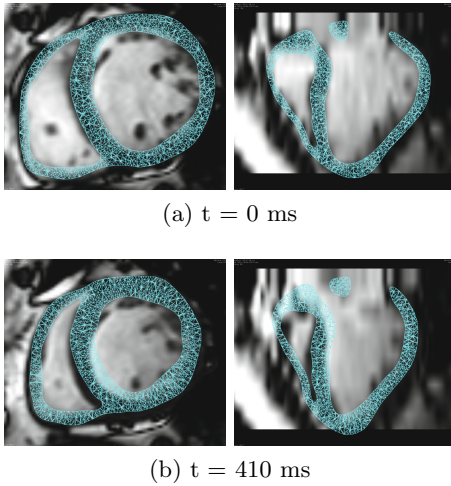
Fig. 5. Simulated mesh for LBBB case 1
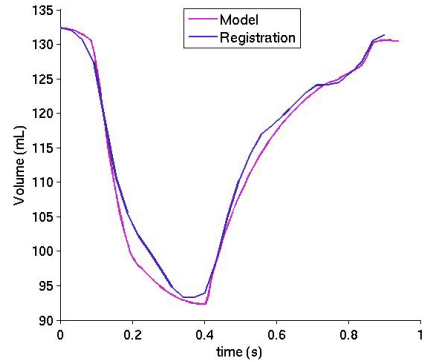
Fig. 6. Comparison between the volume variation computed from the simulation and the one estimated from registered images

the initial simulated volume curve and the one computed after registration gives a relative error of about 25% for both slopes (see Fig. 6), which is of the same order of magnitude as the one after model calibration.

5   Conclusion

In this paper we proposed an innovative calibration method of an electromechanical cardiac model. The model depends on 14 parameters that act on the active, passive and constraint components. The calibration based on the Unscented Transform allowed us to give a fast initialization of 4 or 7 parameters, leaving the others fixed to standard values. The choice of these parameters was made based on a sensitivity analysis on the volume and pressure variation and

confirmed by a Singular-Value-Decomposition analysis. Since the calibration requires only to run several simulations in parallel to estimate these parameters followed by one additional simulated cycle to verify the results, it can easily be used as a preprocessing step before the application of more sophisticated personalization algorithms. Moreover, the calibration performed on the 7 healthy volunteers and 2 heart failure cases allowed us to compare parameters for pathological cases versus healthy controls as a first step toward specificity analysis to classify various pathologies. Additional heart failure cases and observations quantities (global indices of the strain for instance) are required to further improve the calibration and validate this specificity study. Finally, the impact of the calibration on the personalization algorithm also needs to be investigated.

References

1. Bestel, J., Clément, F., Sorine, M.: A Biomechanical Model of Muscle Contraction. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1159–1161. Springer, Heidelberg (2001)
2. Bortone, A.S., Hess, O.M., Chiddo, A., Gaglione, A., Locuratolo, N., Caruso, G., Rizzon, P.: Functional and structural abnormalities in patients with dilated cardiomyopathy. J. Am. Coll. Cardiol. 14, 613–623 (1989)
3. Chapelle, D., Le Tallec, P., Moireau, P., Sorine, M.: An energy-preserving muscle tissue model: formulation and compatible discretizations. Int. J. MultiScale Comput. Eng. (IJMCE) 10(2), 189–211 (2012)
4. Hunter, P., Nash, M., Sands, G.: Computational electromechanics of the heart. In: Computational Biology of the Heart, pp. 345–407 (1997)
5. Huxley, A.: Muscle structure and theories of contraction. Progress in Biophysics and Biophysical Chemistry 7, 255 (1957)
6. Julier, S., Uhlmann, J.: A new extension of the Kalman filter to nonlinear systems. In: Int. Symp. Aerospace/Defense Sensing, Simul. and Controls., vol. 3, p. 26. Citeseer (1997)
7. Liu, H., Shi, P.: Maximum a posteriori strategy for the simultaneous motion and material property estimation of the heart. IEEE Trans. Biomed. Eng. 56(2), 378–389 (2009)
8. Moireau, P., Chapelle, D.: Reduced-order Unscented Kalman Filtering with application to parameter identification in large-dimensional systems. COCV 17, 380–405 (2011)
9. Stergiopulos, N., Westerhof, B., Westerhof, N.: Total arterial inertance as the fourth element of the windkessel model. Am. J. Phys (Heart and Circulatory Phys.) 276(1), H81 (1999)
10. Sundar, H., Davatzikos, C., Biros, G.: Biomechanically-Constrained 4D Estimation of Myocardial Motion. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 257–265. Springer, Heidelberg (2009)
11. Wan, E., Van Der Merwe, R.: The unscented Kalman filter for nonlinear estimation. In: AS-SPCCS, pp. 153–158. IEEE (2000)
12. Xi, J., Lamata, P., Lee, J., Moireau, P., Chapelle, D., Smith, N.: Myocardial transversely isotropic material parameter estimation from in-silico measurements based on a reduced-order unscented Kalman filter. J. Mech. Behav. Biomed. Mater (JMBBM) 4, 1090–1102 (2011)

Temporal Shape Analysis via the Spectral Signature

Elena Bernardis[1], Ender Konukoglu[2], Yangming Ou[1], Dimitris N. Metaxas[3], Benoit Desjardins[1], and Kilian M. Pohl[1]

[1] Dept. of Radiology, University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Microsoft Research, Cambridge, CB3 0FB, UK
[3] Dept. of Computer Science, Rutgers University, Piscataway, NJ 08854, USA

Abstract. In this paper, we adapt spectral signatures for capturing morphological changes over time. Advanced techniques for capturing temporal shape changes frequently rely on first registering the sequence of shapes and then analyzing the corresponding set of high dimensional deformation maps. Instead, we propose a simple encoding motivated by the observation that small shape deformations lead to minor refinements in the spectral signature composed of the eigenvalues of the Laplace operator. The proposed encoding does not require registration, since spectral signatures are invariant to pose changes. We apply our representation to the shapes of the ventricles extracted from 22 cine MR scans of healthy controls and Tetralogy of Fallot patients. We then measure the accuracy score of our encoding by training a linear classifier, which outperforms the same classifier based on volumetric measurements.

1 Introduction

Capturing the shape and function of anatomy through volumetric measurements extracted from 4D medical scans has become of central importance in diagnosing diseases. For example, cardiologists rely on ejection fraction extracted from ultrasound or cine MR scans to assess patients. These volumetric measurements, however, are not sensitive enough to aid the diagnosis of many focal or diffuse cardiac diseases. In this paper, we introduce a new encoding of the shape and its temporal changes based on the spectral signature and show that this encoding is more sensitive for comparing two shapes and their temporal dynamics than volumetric measurements.

Advanced techniques for capturing the changes in shape over time frequently rely on registering the sequence of images and then analyzing the corresponding set of deformation maps [1,2]. Examples of this type of analysis, specific to the heart, include mappings motivated by biomechanical models [3] or mathematical properties [4], from which statistical models of the cardiac ventricles can be extrapolated [5]. While promising, these approaches are susceptible to negatively biasing the analysis due to the underlying assumptions and parameter settings of the registration framework, as well as the accuracy in reducing high dimensional deformation maps to a few features.

An alternative avenue is to characterize the shape of each structure with a low-dimensional set of parameters and use the latter for structure discrimination. Spectral signatures, i.e. eigensystems of the Laplace and Laplace-Beltrami operators, have recently gained popularity as powerful shape descriptors [6,7]. The eigenvalues of the

a: LV & RV segmentations   **b:** Spectrum of the Laplacian   **c:** Learned Eigencurves
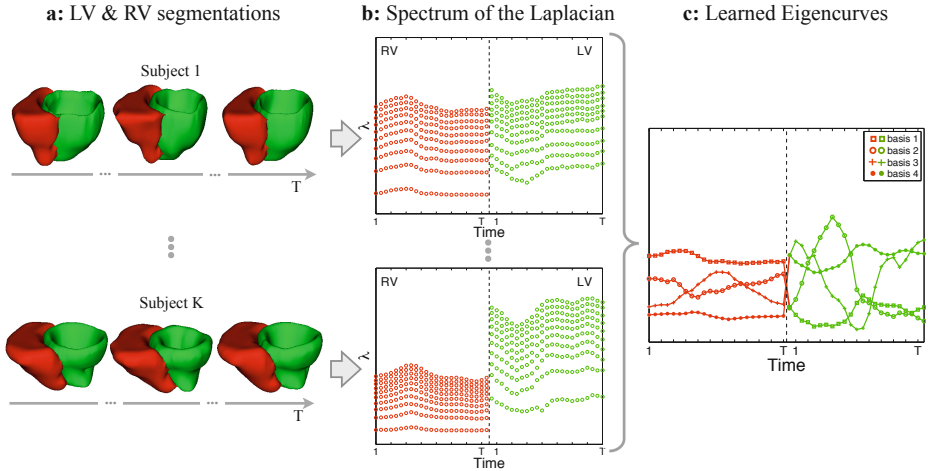
Fig. 1. Our method. **a:** Right (RV) and left (LV) ventricle segmentations for each heart sequence at three sample timepoints. **b:** Eigenvalue curves that encode temporal shape changes displayed as a function of time. **c:** RV and LV reduced shape signatures learned by a low dimensional embedding of these curves, while assuming no prior physical or statistical model of the data.

Laplace operator implicitly carry information on local shape invariants such as curvature, surface area and volume, and allow to encode shape information with few parameters without need of prior registration [8]. In this paper, we specifically explore the adaptation of this technology from capturing individuals shapes to capturing temporal morphologic changes.

Our work is motivated by the observation that small shape deformations lead to minor refinements in the spectral signature. Spectral signatures are thus well suited for capturing temporal shape changes of anatomical structures that vary slowly between measurements, such as cardiac ventricles from cine MR scans. Our representation captures temporal shape changes by first computing the spectral signature for each time point yielding a family of eigenvalue curves (Fig. 1b). We then encode temporal shape changes by a low dimensional embedding of the eigenvalue curves (Fig. 1c). By doing so, our simplistic representation assumes no prior physical or statistical model of the data and only depends on two parameters, namely, the number of eigenvalues and the dimension chosen for the lower dimension embedding.

We apply our representation to the shapes of the ventricles extracted from 22 cine MR scans of 11 healthy controls and 11 Tetralogy of Fallot (TOF) patients. We choose this specific scenario as there is no uncertainty, unlike with other diseases, about the diagnosis, so that the labeling of individual data sets can be viewed as ground truth. In addition, TOF is suitable for temporal analysis as it affects both RV shape and cardiac function. We then measure the accuracy score of our encoding by training a linear classifier and recording the leave-one-out cross-validation accuracy in distinguishing these two populations. Our representation outperforms the linear classifier based on volumetric measurements of the ventricles. Before we describe our encodings and experiments in further detail, we just note that we do not attempt to provide a representation tailored

towards cardiac disease detection but rather derive a shape representation for implicitly encoding temporal morphological changes that can be applied to the cardiac domain.

2   Temporal Shape Encoding

We now present in further detail the extension of the spectral signature of the Laplacian operator to the temporal domain. Laplace operators and their spectra have been studied in mathematics for a long time [9,10,11]. Their introduction in computational shape analysis is, however, rather recent [12]. We start with an introduction to Laplace operators and then describe our spectral shape encoding for temporal shape deformations.

2.1   Spectrum of Laplace Operator

Our brief overview of the Laplace operators is meant to give the necessary background to understand their role in temporal shape encoding. For a more thorough discussion, we refer the reader to [11,12].

We denote an object as a closed bounded domain $\Omega \subset \mathbb{R}^d$ with piecewise smooth boundaries. With respect to medical imaging domain, Ω corresponds to the volume extracted from the segmentation of an anatomical structure. The Laplace operator Δ_Ω on Ω is defined as $\Delta_\Omega f \triangleq \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f$ for a twice differentiable function f, where $x \triangleq \{x_1, \ldots x_d\}$ are the spatial coordinates. The importance of this operator for shape analysis arises from its eigenvalues and eigenfunctions, which are the solutions of the Helmholtz equation with Dirichlet type boundary conditions, $\Delta_\Omega f + \lambda f = 0$, $\forall \mathbf{x} \in \Omega$ and $f(\mathbf{x}) = 0$, $\forall \mathbf{x} \in \partial\Omega$, where $\partial\Omega$ denotes the boundary of the object and $\lambda \in \mathbb{R}$ is a scalar [11]. There are infinite pairs of $\{(\lambda_j, f_j)\}_{j=1}^\infty$ that satisfy this equation and the ordered set of eigenvalues form a positive diverging sequence $0 < \lambda_1 \leq \lambda_2 \leq \ldots$, called the Dirichlet spectrum of Δ_Ω, which we simply refer to as the 'spectrum'.

The spectrum has several advantageous properties for shape analysis in medical image analysis [11]. First, the spectrum encodes information regarding the intrinsic geometry of the object. This information content is due to an identity called *heat-trace* and its equivalent polynomial expansion $Z(\tau) \triangleq \sum_{j=1}^\infty e^{-\lambda_j \tau} = \sum_{m=0}^\infty a_{m/2} \tau^{-d/2+m/2}$, with $\tau > 0$. The coefficients $a_{m/2}$ are given as sums of volume and boundary integrals of some local invariants of Ω [13,11], such as its volume or its surface mean curvature. The functional relation between $\{\lambda_j\}_{j=1}^\infty$ and $\{a_{m/2}\}_{m=0}^\infty$ links the geometry of an object to its spectrum, and is the ingredient that makes the Laplace operator interesting for shape analysis. Second, the eigenvalues are invariant to isometric transformations. This invariance can be even extended to scaling [12], thus eliminating the need to align the shapes to a common coordinate system for further analysis. Finally, the eigenvalues change continuously with the deformations applied to the object's boundary, i.e. there is a continuous link between the differences in eigenvalues and the difference in shape. This continuous link is a critical component for our encoding of temporal shape deformation. For a deeper intuition, let us consider an object that changes its shape with time. We represent the temporal dependence of the object's shape with $\Omega(t)$ and the temporal dependence of its spectra with $\Lambda(t)$. Now, if we interpret the motion of an object as the deformation between two time points, $\Lambda(\cdot)$ not only then captures the geometry of $\Omega(t)$

at a given t but it also provides information regarding the motion of $\Omega(\cdot)$. In this paper, we make use of both information when encoding 3D+t(ime) objects.

2.2   Learning Temporal Shape Changes

We now outline the framework for encoding 3D + t. Given a set of K 3D+t volumes defined over T time points, we first compute for each volume $i \in \{1, \ldots, K\}$ and time point $t \in \{1, \ldots, T\}$ the first N eigenvalues of the spectrum (see also Fig. 1b), which we denote with $\{\lambda_1^i(t), \lambda_2^i(t), \ldots, \lambda_N^i(t)\}$. We then construct temporal curves from the eigenvalues across the N signatures of each volume, i.e. $\lambda_j^i := \{\lambda_j^i(1), \ldots, \lambda_j^i(T)\}$, for $j \in \{1, \ldots, N\}$, is the jth temporal (eigenvalue) curve of volume i. The final step encodes the temporal shape deformations by learning a low dimensional embedding of those curves across all K volumes. By doing so, further analysis of our encoding always has to consider the subject specific matrix of eigenvalues defining curves across time instead of viewing λ_j^i individually.

For dimensionality reduction, we apply Non-negative Matrix Factorization (NMF) [14] to the $N \cdot K$ temporal curves. In general, the dimension of the data matrix V is $M \times O$ where M is the number of measurements and O the number of objects. In our temporal shape encoding, each column of the data matrix V represents a temporal curve λ_j^i. Thus, the dimension of V becomes the (number of timepoints T) $\times (K \cdot N)$. NMF then factorizes matrix V into a basis matrix H and coefficient matrix W so that

$$V \approx WH \text{ subject to minimizing } F(W, H) \triangleq \sum_{i=1}^{N \cdot K} \sum_{\mu=1}^{T} [V_{i\mu} \log(WH)_{i\mu} - (WH)_{i\mu}],$$

where V refers to the entries of the corresponding matrix. Setting the number of basis vectors to B and $b \in \{1, \ldots, B\}$, the optimal H and W are obtained by finding the local minimum of $F(\cdot, \cdot)$ via the following iterative algorithm:

$$\textbf{1: } W_{ib}' \leftarrow W_{ib} \sum_{\mu=1}^{T} \frac{V_{i\mu}}{(WH)_{i\mu}} H_{b\mu} \quad \textbf{2: } W_{ib} \leftarrow \frac{W_{ib}'}{\sum_{j=1}^{N \cdot K} W_{jb}'} \quad \textbf{3: } H_{b\mu} \leftarrow H_{b\mu} \sum_{i=1}^{N \cdot K} W_{ib} \frac{V_{i\mu}}{(WH)_{i\mu}}$$

All temporal curves can now be represented by the basis matrix H and the corresponding B coefficients provided by W. These B coefficients across all the sequences of N eigenvalues are then the temporal shape encodings of our 3D+t objects.

We note that one could have chosen any other dimensionality reduction method. We simply choose NMF as it does not make any assumption of the underlying distribution, unlike for example Principle Component Analysis as motivated by our experiments.

3   Experimental Setup

We evaluate our temporal shape encoding by applying it on ventricle segmentations from short-axis cardiac MRI scans and comparing its accuracy score in a classification problem to discriminate healthy individuals from patients who had Tetralogy of Fallot (TOF) corrected surgery in infancy. Our dataset includes the cine MR scans of 11 TOF cases and 11 healthy volunteers (K=22). TOF patients are post repair and age-matched

to normal controls. The cine MR scans of each case are normalized to 23 timepoints. A medical expert then semi-automatically segmented the blood pool of the right ventricle and myocardium of the left ventricle at the end-diastole (ED) timepoint using 'Segment' [15] with manual corrections of the results. For simplicity, we refer to the blood pool of the right ventricle just as right ventricle or RV, and to the myocardium of the left ventricle as left ventricle or LV. We then propagate this segmentation to the other time points via the non-rigid registration of [16]. We refine those segmentations via dilation and erosion to remove possible holes, disconnections or topology changes caused by registration errors. Sample segmentation sequences are shown in Fig. 1a.

For computing the Laplace spectrum, we define the shapes of LV and RV based on the corresponding segmentations. We then compute the Dirichlet eigenvalues using finite differences on the regular grid. While using a high order FEM method might provide more accurate eigenvalues, this is outside the scope of this article as our focus is to apply the signature to the temporal domain, hence we use the simplest implementation. Furthermore, one can also consider using Neumann eigenvalues however they would be more sensitive to noisy segmentations e.g. isolated islands.

We analyze the accuracy of our new temporal shape encoding in describing the shape changes of the RV and LV by first applying it as well as related representations to our sequences of segmentations. We then measure the accuracy of a linear Support Vector Machine (SVM) [17] in correctly labeling images based on those features via the leave-one-out principle. Table 1 records the outcome when applying the classifier to LV and RV separately as well as when we combine the eigencurves of both ventricles.

We start by extracting the volumes and computing the first eigenvalue of the RV and LV. We view each volume or first eigenvalue as a feature so that classification of every RV and LV is based on 23 features (given from the original 23 timepoints). When concatenating the ventricles (which we will denote by RV-LV), classification is based on 46 features instead. Classification results based on volume and based on the first eigenvalue are shown in Table 1 (a) and (b) respectively. Simply using volume for distinguishing shape changes between TOF and normal subjects only leads to an accuracy of approximately 60% for LV and RV separately and of 68% when combined. As mentioned in the previous section, the functional relation between $\{\lambda_j\}_{j=1}^{\infty}$ and $\{a_{m/2}\}_{m=0}^{\infty}$ is the link between the geometry of an object and its spectrum. In particular, the first eigenvalue is directly proportional to the volume of object i since $a_0 = (4\pi)^{-3/2} V_i$. As predicted, the results obtained using the first eigenvalue alone are comparable to the volume ones.

As shown in the rest of the experiments, increasing the number of eigenvalues adds shape information, improving the overall recall. We start by considering the eigenvalue sequences λ_j^i individually. We do so by computing the eigenvalues for each of the hearts, setting the spectrum cutoff at $N = 100$. For comparison to our new encoding, the results of Table 1 (c) do not explicitly model temporal dependencies as we first combine the $23 \cdot N$ features for each ventricle (or $46 \cdot N$ for RV-LV). We then reduce the $M = 23 \cdot N$ eigenvalues (or $46 \cdot N$ for RV-LV) for the $O = K$ hearts via NMF (see Sec. 2.2) to $B = 22$ features. Parameters N and B are chosen empirically to be minimum while maximizing classification accuracy. We experimented for N from 1 to 200 and for B from 2 to 200. Furthermore, B is set to the same for the LV, RV, and RV-LV scenarios. The resulting classification scores (c) are higher compared to (a,b).

Table 1. TOF classification results. For each experimental setup, we record precision (prec.), recall and accuracy for LV, RV and RV-LV, obtained from our leave-one-out cross validation. The first eigenvalue (b) alone performs similarly to the volume feature (a). Increasing number of eigenvalues (c) improves the performance, especially recall. d-e) Our 3D+t shape signatures method yields increased scores. f) Choosing PCA instead of NMF on the same eigencurves used for (d) results in lower scores.

	LV			RV			RV-LV		
	Prec.	Recall	Accuracy	Prec.	Recall	Accuracy	Prec.	Recall	Accuracy
a: volume	0.7727	0.5862	0.6136	0.9394	0.5741	0.6212	0.9301	0.6274	0.6888
b: λ_1	0.8182	0.5625	0.5909	0.8636	0.6441	0.6932	0.8864	0.6290	0.6818
c: $\lambda_{1:1:100}$	0.7143	0.5914	0.6104	0.7557	0.6425	0.6676	0.8182	0.7674	0.7851
d: $\lambda_{1:1:100}$ (3D+t)	0.6970	0.8214	0.7727	0.7576	0.8065	0.7879	0.8788	0.8286	0.8485
e: $\lambda_{1:10:100}$ (3D+t)	0.7879	0.8667	0.8333	0.8182	0.8438	0.8333	0.9545	0.8400	0.8864
f: PCA $\lambda_{1:1:100}$	0.7273	0.6957	0.7045	0.7273	0.7500	0.7424	0.7879	0.7647	0.7727

We further improve the classification score by using our new 3D+t temporal shape representation of Sec. 2.2 (Table 1 (d,e)). Here, we reduce the dimensionality of the $N \cdot K$ eigencurves from $T = 23$ entries (respectively 46 for the RV-LV) to $B = 4$ after exploring the entire range from 2 to 23. Fig. 2a illustrates sample eigencurves. We then feed $B \cdot C$ features to the classifier, where C is the number of curves selected for the classification step. In line (d), we use all the eigencurves for classification, hence setting $C = N (= 100)$, while we subsample the eigencurves in (e) setting $C = 10$. The 3D+t encoding yields a higher precision and significantly improves recall, hence giving an overall improved accuracy. As shown from the classification results in (e), selecting a subset of the eigencurves, once all the eigencurves have already been used for the learning step, allows to improve the outcome of the classifier, indicating that another method for data compression or classification could improve even further the results of our temporal shape descriptor as well as that of the other scores.

To motivate the use of NMF, we also applied PCA (f) to learn the temporal eigencurves using the same setup of (d). The accuracy scores drop to the level we measured for the volumes scores indicating that the Gaussian assumption of PCA is violated by these sets of temporal eigencurves. Our quantitative findings are also reflected in the visual comparison of the basis vectors for $B = 4$ of the two dimensionality reduction methods in Fig. 2b. While the bases obtained by PCA are very noisy, the NMF ones are cleaner and better describe the smooth temporal changes observed in the eigencurves.

Summary. In this paper, we exploit the implicit local shape properties captured by spectral signatures, i.e. eigensystems of the Laplace and Laplace-Beltrami operators, and adapt them to capture morphological changes over time. We propose a fairly simple encoding based on the observation that small shape deformations lead to minor refinements in the spectral signature. The shape analysis is independent of the original segmentation used and, given initial segmentations, computing the spectral signature is independent of registration. The accuracy obtained in the classifications demonstrate that that our temporal-shape representation can be successfully used to classify TOF cardiac disease patients. Our results indicate that our new temporal shape representation better incor-

a: Sample eigencurves for healthy (left) and TOF (right) cases
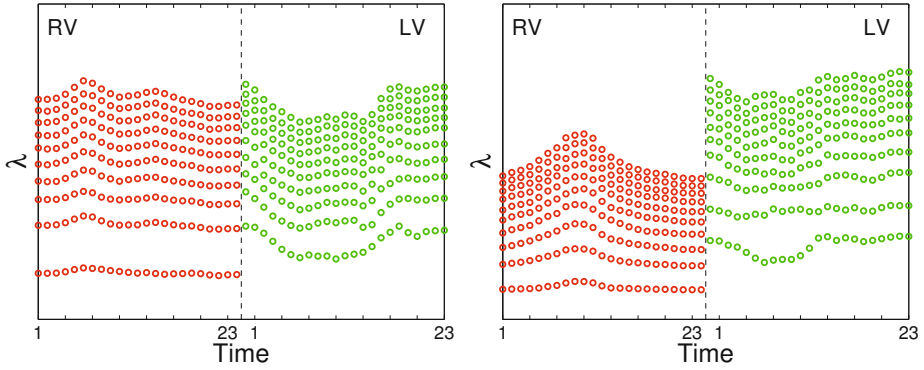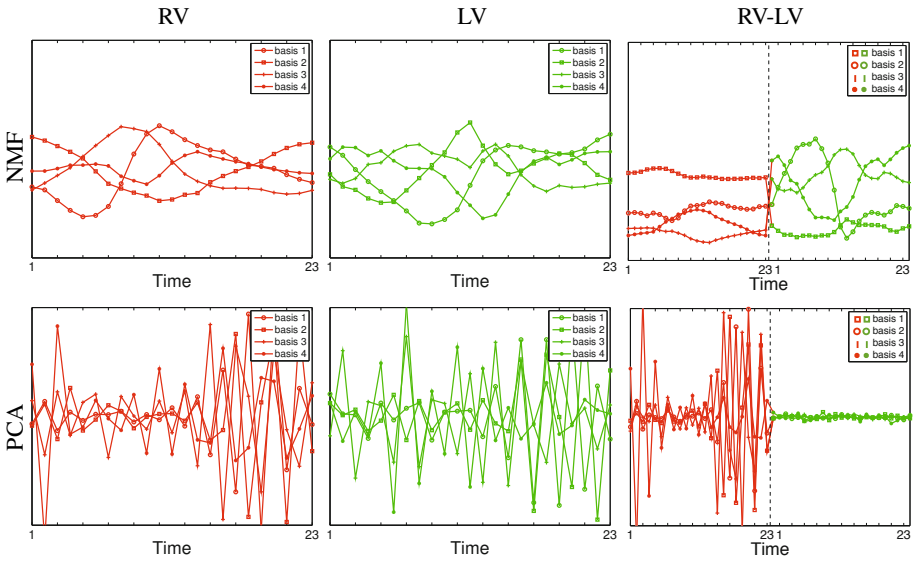
b: Reduced spectral signatures

Fig. 2. Shape signatures. **a:** Sample eigencurves before reduction. **b:** Learned spectral signatures reduced via NMF and PCA for RV and LV (over 23 timepoints), and for RV-LV (over 46 timepoints). In the absence of the correct initial statistical model, PCA results in a high level of noise and does not capture the smooth temporal changes of the eigencurves.

porates the temporal relation of the data. Thus, we are able to better capture RV and LV deformations in this population than scores capturing shape changes over time by separately measuring the 3D shape at each timepoint.

References

1. Ardekani, S., Weiss, R.G., Lardo, A.C., George, R.T., Lima, J.A.C., Wu, K.C., Miller, M.I., Winslow, R.L., Younes, L.: Cardiac motion analysis in ischemic and non-ischemic cardiomyopathy using parallel transport. In: ISBI, pp. 899–902 (2009)
2. Beg, M.F., Helm, P.A., Mcveigh, E., Miller, M.I., Winslow, R.L.: Computational cardiac anatomy using MRI. Magnetic Resonance in Medicine (2004)
3. Wang, X., Chen, T., Zhang, S., Metaxas, D., Axel, L.: LV Motion and Strain Computation from tMRI Based on Meshless Deformable Models. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 636–644. Springer, Heidelberg (2008)
4. Helm, P., Beg, M.F., Miller, M.I., Winslow, R.L.: Measuring and mapping cardiac fiber and laminar architecture using diffusion tensor mr imaging. Annals of the New York Academy of Sciences 1047(1), 296–307 (2005)
5. Mansi, T., Durrleman, S., Bernhardt, B., Sermesant, M., Delingette, H., Voigt, I., Lurz, P., Taylor, A.M., Blanc, J., Boudjemline, Y., Pennec, X., Ayache, N.: A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 214–221. Springer, Heidelberg (2009)
6. Reuter, M., Niethammer, M., Wolter, F.-E., Bouix, S., Shenton, M.: Global medical shape analysis using the volumetric laplace spectrum. In: Proceedings of the 2007 International Conference on Cyberworlds, NASA-GEM Workshop, pp. 417–426. IEEE Computer Society, Los Alamitos (2007)
7. Reuter, M.: Laplace Spectra for Shape Recognition. Books on Demand GmbH (2006)
8. Niethammer, M., Reuter, M., Wolter, F.-E., Bouix, S., Peinecke, N., Koo, M.-S., Shenton, M.E.: Global Medical Shape Analysis Using the Laplace-Beltrami Spectrum. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 850–857. Springer, Heidelberg (2007)
9. Weyl, H.: Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen. Math Ann., 441–469 (1912)
10. Kac, M.: Can one hear the shape of a drum? The American Mathematical Monthly 73(7), 1–23 (1966)
11. Courant, R., Hilbert, D.: Method of Mathematical Physics, vol. I. Interscience Publishers (1966)
12. Reuter, M., Wolter, F.E., Peinecke, N.: Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. Computer-Aided Design 38, 342–366 (2006)
13. Protter, M.: Can one hear the shape of a drum? Revisited. SIAM Review 29(2), 185–197 (1987)
14. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS, pp. 556–562 (2000)
15. Heiberg, E., Sjogren, J., Ugander, M., Carlsson, M., Engblom, H., Arheden, H.: Design and validation of segment - freely available software for cardiovascular image analysis. BMC Medical Imaging 10(1), 1 (2010)
16. Ou, Y., Sotiras, A., Paragios, N., Davatzikos, C.: Dramms: Deformable registration via attribute matching and mutual-saliency weighting. Medical Image Analysis 15(4), 622–639 (2011)
17. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 2, 121–167 (1998)

Joint T1 and Brain Fiber Log-Demons Registration Using Currents to Model Geometry

Viviana Siless[1,2], Joan Glaunès[3], Pamela Guevara[4], Jean-François Mangin[2], Cyril Poupon[2], Denis Le Bihan[2], Bertrand Thirion[1,2], and Pierre Fillard[1,2]

[1] Parietal Team, INRIA Saclay-Île-de-France, Saclay, France
viviana.siless@inria.fr
http://parietal.saclay.inria.fr
[2] CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France
[3] MAP5, CNRS UMR 8145, Université Paris Descartes, 75006 Paris, France
[4] University of Concepción, Concepción, Chile

Abstract. We present an extension of the diffeomorphic Geometric Demons algorithm which combines the iconic registration with geometric constraints. Our algorithm works in the log-domain space, so that one can efficiently compute the deformation field of the geometry. We represent the shape of objects of interest in the space of currents which is sensitive to both location and geometric structure of objects. Currents provides a distance between geometric structures that can be defined without specifying explicit point-to-point correspondences. We demonstrate this framework by registering simultaneously T_1 images and 65 fiber bundles consistently extracted in 12 subjects and compare it against non-linear T_1, tensor, and multi-modal T_1+ Fractional Anisotropy (FA) registration algorithms. Results show the superiority of the Log-domain Geometric Demons over their purely iconic counterparts.

Keywords: Registration, neural fibers, diffeomorphism, Demons Algorithm, intensity-base registration, tensor-base registration, log-domain.

1 Introduction

Non-linear image registration is one of the most challenging tasks in medical image analysis. For inter-individual comparison, registration should align images as well as cortical and internal structures such as sulcal lines and fibers. Non-linear registration algorithms can be categorized into iconic, geometric and hybrid.

Iconic, or image-based registration [13,3,10] finds a voxel-wise mapping between a source and a target image. Schematically, iconic registration is mainly driven by the contours of the image, and without prior knowledge, it is difficult to coregister regions with little contrast. For instance, brain white matter appears uniformly white in T_1 images, giving no relevant information to the iconic registration, while it is composed of neural fibers connecting cortical areas. Diffusion Tensor Imaging (DTI) can be used to reveal the microscopic structure of the white matter. Tensor-based registration was recently proposed to improve

white matter alignment [17,15]. However, misregistration may persist in regions where the tensor field appears uniform, as shown in [4].

Geometric registration specifically targets the alignment of Structures of Interest (SOI), such as in [16] for cortical surfaces, or [4] for fiber bundles. While those clearly improve SOI registration, they are in general not suitable for aligning other structures than those used specifically during registration.

Hybrid techniques propose to jointly consider SOI and images during registration. For instance, [1,8] used the mathematical framework of measures and currents to simultaneously register images and geometric descriptors, while [12] proposed a Markovian solution to the same problem.

We present an hybrid registration algorithm based on the efficient framework of Demons, where we combine iconic and geometric registration. The geometry is represented in the space of currents which provides a metric sensitive to shape.

The rest of the paper is organized as follows. First, we propose a mathematically sound extension of the Geometric Demons(GD), the Log-domain Geometric Demons (LGD), that relies in the log-domain framework for computation and handles geometric constraints in the space of currents. Then, we evaluate the LGD with bundles constraints on a dataset of 12 subjects and compare them with a scalar [14], a tensor [15], and Ants [2] a multi-modal registration.

2 The Log-Domain Geometric Demons

2.1 The Diffeomorphic Demons

In image registration we search for a displacement field s between a fixed F and moving M image, that maps as accurately as possible corresponding structures in both images. Ideally the displacement field s minimizes the distance between the fixed and the moving image, while holding some properties such as being diffeomorphic. In the Demons framework[13] a correspondence field c was introduced to make the minimization of the functional energy tractable: $E(c,s) = \frac{1}{\sigma_i^2}\text{Sim}(F, M \circ c) + \frac{1}{\sigma_x^2}\text{dist}(s,c)^2 + \frac{1}{\sigma_T^2}\text{Reg}(s)$ where Sim is a similarity measure between images defined by the sum of square differences (SSD) and Reg a regularization term chosen to be an harmonic energy. The amount of regularization is controlled with σ_T while σ_i accounts for the image noise. The term $\text{dist}(s,c)^2$ imposes the displacement field s to be close to the correspondence field c. σ_x weights the spatial uncertainty on the deformation. The energy minimization is performed by alternating minimization w.r.t. c and s. In [14], small deformations parametrized by a dense displacement field u are used: $c \leftarrow s \circ \exp(u)$, $\exp()$ being the exponential map in the Lie group sense, which ensures that the result is diffeomorphic.

2.2 Geometric Demons

To incorporate geometric constraints in the Demons framework, in [11] the definition of c was extended to carry information from both image and geometry.

Let us denote by $\mathcal{G}^F$ (resp. $\mathcal{G}^M$) the fixed (resp. moving) geometric descriptors. Therefore a new energy was defined:

$$E(c, s) = \frac{1}{\sigma_i^2} \left[\text{Sim}_I(F, M \circ c) + \text{Sim}_G(c \star \mathcal{G}^F, \mathcal{G}^M) \right] +$$
$$\frac{1}{\sigma_x^2} \text{dist}(s, c)^2 + \frac{1}{\sigma_T} \text{Reg}(s), \qquad (1)$$

where Sim_I is the image similarity measure, Sim_G the geometry similarity measure, and $c \star \mathcal{G}^F$ denotes the action of c on the geometry.

Following [14], c was parametrized by an update field, as the additive combination of an image update field u_I and a geometric update field u_G. Non-intersecting domains were defined, as ideally, one should use u_G only where geometric information is relevant and use u_I elsewhere. In the case of fibers, the geometric domain will remain within the white matter. Thus, let Ω_G be the definition domain of u_G (where geometry is defined), and the definition domain of u_I be $\Omega_I = \Omega - \Omega_G$. Then we can define $c = \exp(u_I + u_G)$. The following relationships hold: $c \star \mathcal{G}^F = \exp(u_G) \star \mathcal{G}^F$ and $M \circ c = M \circ \exp(u_I)$.

Geometric Demons incorporates the following energy to calculate u_G:

$$E_G(s, u_G) = \frac{1}{\sigma_i^2} \text{Sim}_G(s \circ \exp(u_G) \star \mathcal{G}^F, \mathcal{G}^M) + \frac{1}{\sigma_x^2} \int_{\Omega_G} \|u_G\|^2, \qquad (2)$$

Being s the deformation field from F to M. Thus, the inverse of the s gives the geometric deformation. In section 2.3 we show an efficient approximation for obtaining the geometric deformation in the log-domain space.

2.3   Log-Domain Geometric Demons

The log-domain demons avoids the inversion of the deformation field by redefining s with the exponential map: $s = \exp(v)$. Then $s \circ \exp(u) = \exp(v) \circ \exp(u)$ and the Baker-Campbell-Hausdorff(BCH) formula yields $\log(\exp(v) \circ \exp(u)) \approx v + u + 1/2[v, u] + 1/12[v, [v, u]] + ...$ where $[v, u]$ is the Lie bracket.

Then, the LogGeometric Demons algorithm is defined as follows:

1. Choose a starting spatial transformation $s = \exp(v)$
2. Given s, u_I, compute the update field u_I as in [14]
3. Given s, u_G, compute the update field u_G by minimizing Eq. (2)
4. Let $u \leftarrow s \circ \exp(u_I + u_G)$
5. $v \leftarrow \log(\exp(v) \circ \exp(u))$ using BCH approximation and $\exp(u)$ is efficiently computed with a few compositions, look [14] for further details.
6. let $v \leftarrow K_{diff} \star v$ where K_{diff} is a Gaussian convolution kernel
7. $s = \exp(v)$ and $s^{-1} = \exp(-v)$
8. Go to 2. until convergence

With this new definition we can efficiently compute the geometric deformation.

Calculation of u_G in the Space of Currents. In GD the closest point distance was proposed as it does not need explicit correspondences between points and it can be a good approach for comparing single fiber bundles representatives as in [11]. In contrast, by representing geometry in the space of currents, we have a pose and shape-sensitive measure which permits to define a distance between bundles containing different number of fibers.

Let $\mathcal{G}$ be a set of continuous curves. We define the sequence of discretized points in $\mathcal{G}$ as $\mathcal{G} = (x_1, ..., x_N)$, N being the number of points. We can associate to this sequence a specific measure given by the vector valued Diracs: $\mu_{\mathcal{G}} = \sum_{i=1}^{n-1} \tau_{\mathcal{G},i} \delta_{c_{\mathcal{G},i}}$ where $c_{\mathcal{G},i} = (x_i + x_{i+1})/2$(center point) and $\tau_{\mathcal{G},i} = x_{i+1} - x_i$ (tangent vector) if x_i and x_{i+1} belong to the same curve, otherwise $\tau_{\mathcal{G},i} = 0$.

Following [6], let W be a reproducible kernel Hilbert space (r.k.h.s) of vector fields with kernel K_β^W isotropic and Gaussian of size β: vector fields in W are convolutions between any square-integrable vector fields and the convolution square root of the kernel. Then, the vector space of currents is a dense span of the set of all the vector valued Diracs currents $\tau \delta_c$ for any $\tau, c \in \mathbb{R}^3$. A Dirac current may be seen as an oriented segment entirely concentrated at point c. The scalar product between two sums of vector valued Diracs expresses conveniently in terms of the kernel K_β^W:

$$\langle \mu, \mu' \rangle = \langle \sum_{i=1}^{N} \tau_i \delta_{c_i}, \sum_{j=1}^{M} \tau'_j \delta_{c'_j} \rangle = \sum_{i=1}^{N} \sum_{j=1}^{M} K_\beta^W(c_i, c'_j) \tau_i . \tau'_j \tag{3}$$

Having a fixed and a moving geometric descriptor $\mathcal{G}^F = (x_1, ..., x_N)$ and $\mathcal{G}^M = (y_1, ..., y_M)$ N, M being the number of points, the distance is defined as follows:

$$d^2(\mathcal{G}^F, \mathcal{G}^M) = ||\mathcal{G}^F - \mathcal{G}^M||_{W^*}^2 = ||\mathcal{G}^F||_{W^*}^2 + ||\mathcal{G}^M||_{W^*}^2 - 2\langle \mathcal{G}^F, \mathcal{G}^M \rangle_{W^*} \tag{4}$$

The distance measures geometrical differences both in pose and shape. With β we define the kernel size, and points at distances much larger than β have a large distance disregarding the shape. Also, when distances are much smaller than β, they are taken as noise, and thanks to the smoothing effect of the kernel they are not taken into account. So the distance captures first misalignment and then shape dissimilarities until a noise level quantified by β is reached.

Let us define the action of the correspondence field c on $\mathcal{G}$ as: $c \star \mathcal{G} = \{s \circ \exp(u_G)(x_i)\}_{i \in [1,N]} \approx \{s(x_i) + u_G(x_i)\}_{i \in [1,N]}$. Since we are dealing with discrete points, we choose to parametrize the dense update field u_G by a finite set of vectors $u_{G,i}$ using radial basis function interpolation: $u_G(x) = \sum_{i=1}^{N} h(\|x - x_i\|)\lambda_i$. When $h(x) = e^{-(r)^2}$, λ_i are calculated such that $u_G(x_i) = u_{G,i} \forall i$. Let us define the matrix A such that $[A]_{i,j} = h(\|x_i - x_j\|)$ ($[A]_{i,j}$ denotes the (i,j) entry of A), $H(x)$ the vector such that $[H(x)]_i = h(\|x - x_i\|)$ and $U = [u_{G,1}, ..., u_{G,N}]$. We can write: $u_G(x) = H(x)A^{-1}U$. Solving $\nabla E_G(s, u_G) = 0$ w.r.t. u_G narrows down to optimization for the $u_{G,i}, \forall i$. After differentiation, we obtain:

$$u_{G,i} = \frac{\nabla_{u_{G,i}} ||c \star \mathcal{G}^F - \mathcal{G}^M||_{W^*}^2}{1 + \frac{\sigma_i^2}{\sigma_x^2}[H(s(x_i))A^{-1}]_i} \tag{5}$$

Defining Ω_G for Bundles. Since we want fibers to influence the deformation near the definition domain, we define the domain as the union of $\gamma-$radius balls B centered at each coordinate x_i. We control the influence by varying γ and thus, dilating the domain. We define a binary map $\Omega_G^\gamma = \bigcup_{i=1}^{N} B(x_i, \gamma)$. The domain of the image correspondence field is the complementary of Ω_G^γ: $\Omega_I^\gamma = \Omega \backslash \Omega_G^\gamma$.

3   Joint T_1 MRI and Brain Bundle Registration

3.1   Data Description

We used the NMR database of 12 healthy volunteers scanned with T_1 ($256 \times 256 \times 124, .9375 \times .9375 \times 1.2mm$) and DW-MRI ($128 \times 128 \times 60, 1.875 \times 1.875 \times 2mm$) [9]. 200 encoding gradients were used for the diffusion sequence. Using [7], we obtained corresponding fiber bundles between several subjects. 100 bundles were consistently identified in all subjects. The longest 65 bundles distributed in both hemispheres were retained for the experiments. For each subject we obtained the linear transformation from B_0 to T_1 to align bundles with T_1 images.

3.2   Experiments

Two experiments were conducted. First, we exhaustively analyze the parameter γ in ω_G^γ defined in Sec. 2.3 to understand its effect on registration accuracy. Second, we compared the performance between Symmetric Log Domain Demons (SLDD), the Symmetric Tensor Demons (STD) and Ants. The inverted deformation field was applied to the fibers to display the registration. Each algorithm was tested on the 3-steps multi-scale approach with 15, 10 and 5 iterations at each scale (from small to large). We set the currents kernel size β by using a robust estimator of the maximum distance between bundles thresholded at 20mm.

Influence of γ. In the first experiment, the 11 subjects were registered onto one, arbitrary chosen as the target subject. We varied γ from 0 to 4.5, where γ is scaled by the smallest voxel size. We divided our bundles in 5 sets (13 bundles each, with bundles of ~ 3 fibers, each of 21 points), and used jointly 4 to train, and the left one to test. The following results show the average of the 5 possible permutations of choosing the test set. We show results over training set (a), test set (b), and the image (c) for the increasing values of γ in Fig. 1. As expected, fiber matching improves as γ increases(a) at the expense of image alignment(c). Indeed, when fibers have a large influence on their neighborhood, image-driven forces are discarded, leading to poor image registration. Also, comparing (a) and (b) we note that $\gamma = 4.5$ is overfitting the fibers, misleading the overall registration. $\gamma = 3.0$ largely improves fiber alignment, while keeping a good match between images. In the sequel, a $\gamma = 3.0$ will be used. In some cases image matching is slightly improved when using $\gamma = 1.5$ pointing out that geometry may indeed help image registration to avoid local minima.
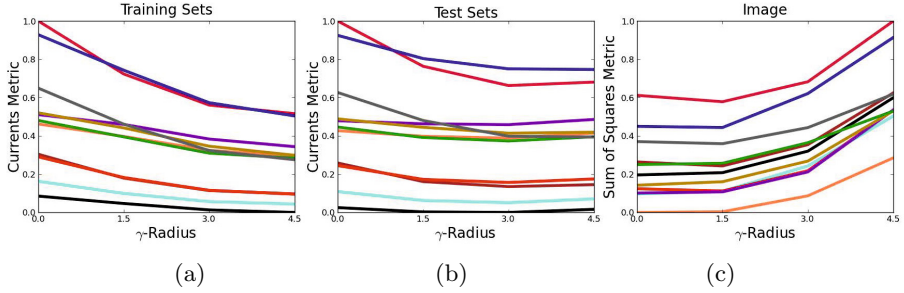
Fig. 1. Influence of γ. Similarity measure average of the 5 training set permutations with varying γ. Different color curves represent the 11 subjects.

Comparison with SLDD, STD and ANTS. For this experiment we register all subjects to each subject for each permutation of the training set. For SLDD, we registered 11 T_1 images onto the 12 permuted target and applied the inverted deformation field to the bundles. For STD, we extracted tensors using [5] and registered each of them onto the 12 permuted target tensor image. Then, inverted deformation fields were applied to each subject's fibers in the DWI space. Finally, the linear transformation calculated between the target B_0 and T_1 images was used to carry fibers to the T_1 space. For Ants, we extracted the FA from the tensors obtained using [5], and aligned them to their T_1 image. We use the cross correlation setting with weighting equally image and FA. Then affine and non-rigid resulting transformation were applied to images, while the inverse of the non-rigid and the affine were applied to the fibers. We show the average metric over training sets (i), test sets (j) and image (k) of registering all subjects to each one with the methods mentioned above in Fig. 2.

As expected, LGD further improved fiber registration in (i) compared to the other algorithms. However, the training set contains the fibers used during registration; we explicitly optimize a metric evaluated on those fibers. Analyzing the results over the test set in (j) we see a similar performance compared to Ants, and STD, which is remarkable as STD is using information from tensors over the whole dense grid, and Ants does a cross correlation between the whole FA grid and T_1. By contrast, in LGD the deformation field was obtained using only sparse information coming from selected fibers, which are not defined in the regions tested in (j). Therefore having a similar performance is very promising. We can also see in (k) the image registration for STD was extremely poor. We time all algorithms with an Intel Xeon 8proc. 2.53GHz, 11.8Gb and obtained: SLDD=19.61min, STD=10.75min, Ants=25.63min, and LGD=12.51min.

4   Discussion

We compared our algorithm against a scalar (SLDD), a tensor (STD), and a multi-modal (Ants) registration. Results show that bundle alignment was highly improved comparing to other algorithms. We get accurate results even for testing

set fibers where no information was used from the support regions of the those fibers, and STD and Ants were using information from the whole grid. This shows that a small set of fibers might be sufficient for a proper registration of the white matter across subjects. Moreover, while fiber alignment is improved, the efficiency of the image alignment is maintained. When evaluating the algorithm for the different γ values, we could see that $\gamma = 1.5$ better registers missing structures than $\gamma = 0$ and than SLDD. However, we believe there is a trade-off to make between image and fiber alignment, and $\gamma = 3$ notably improved fiber alignment while still obtaining good image registration results. By using labeled

(d) Original   (e) LGD   (f) SLDD   (g) STD   (h) Ants

(i)   (j)   (k)
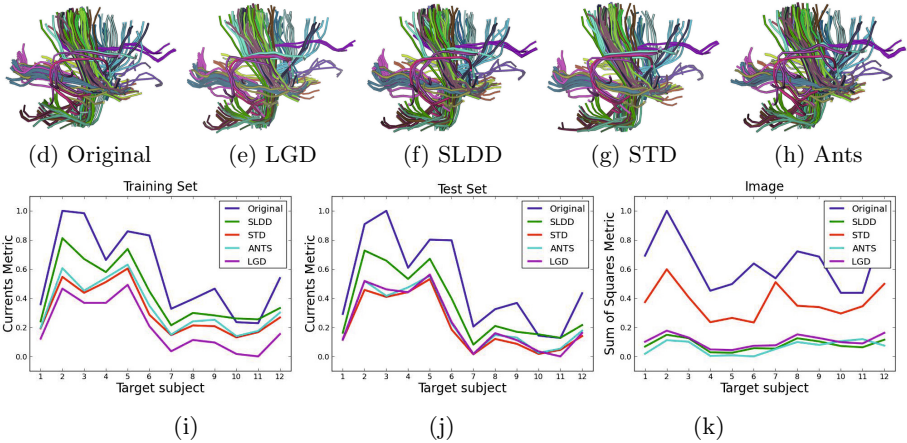
Fig. 2. Comparison of SLDD, STD, Ants and LGD   Top: Target fibers overlapped with registered fibers from an arbitrary chosen registration. 29 fibers were arbitrary preselected for clarity. Corresponding fibers in subjects share colors. **Bottom:** Average of the metrics obtained from registering 11 subjects to the target subject.

bundles instead of purely tensor information, we add relevant features that were previously extracted as prior such as region connection. Nevertheless, the efficacy of trusting bundles is open to discussion, and bundles classification is an active topic in research, so we believe this information should not be discarded.

5   Conclusion

We extended the GD algorithm to the log-domain space, and combined it with currents to compare our geometric structures and respect their shape. Our algorithm perfomed similarly as other competitive algorithms on image and showed a large improvement in fiber alignment. A unique mapping for images and geometric structures is obtained, giving a consistent framework for analyzing and comparing results between voxel-based morphometry and shape of SOI.

Having disjoint domains for geometry and iconic may potentially limit the incorporation of new geometric structures such as sulcal lines or features. This

choice was done to guarantee a closed-form solution when zeroing out the gradient of the criterion. Moreover, both update fields are eventually combined and smoothed out, lowering the effect of the disjoint domains. Future work will further address the remaining mathematical issues to allow overlapping domains.

Based on the efficiency of the algorithm, it would be interesting to combine it with the clustering of fibers: increased fiber registration can help clustering algorithms, which can in turn guide the registration.

References

1. Auzias, G., et al.: Diffeomorphic brain registration under exhaustive sulcal constraints. IEEE Trans. Med. Imaging (January 2011)
2. Avants, B.B., et al.: A reproducible evaluation of ants similarity metric performance in brain image registration. NeuroImage 54(3), 2033–2044 (2011)
3. Beg, M.F., et al.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61(2), 139–157 (2005)
4. Durrleman, S., et al.: Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. NeuroImage 55(3), 1073–1090 (2011)
5. Fillard, P., et al.: Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. IEEE Trans. Med. Imaging 26(11), 1472–1482 (2007)
6. Glaunès, J., et al.: Large deformation diffeomorphic metric curve mapping. IJCV 80, 317–336 (2008), 10.1007/s11263-008-0141-9
7. Guevara, P., et al.: Robust clustering of massive tractography datasets. Neuroimage 54(3), 1975–1993 (2011)
8. Ha, L., Prastawa, M., Gerig, G., Gilmore, J.H., Silva, C.T., Joshi, S.: Image Registration Driven by Combined Probabilistic and Geometric Descriptors. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 602–609. Springer, Heidelberg (2010)
9. Poupon, C., et al.: A database dedicated to anatomo-functional study of human brain connectivity. In: 12th HBM Neuroimage, Florence, Italie, vol. 646 (2006)
10. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic Registration Using B-Splines. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
11. Siless, V., Guevara, P., Pennec, X., Fillard, P.: Joint T1 and Brain Fiber Diffeomorphic Registration Using the Demons. In: Liu, T., Shen, D., Ibanez, L., Tao, X. (eds.) MBIA 2011. LNCS, vol. 7012, pp. 10–18. Springer, Heidelberg (2011)
12. Sotiras, A., Ou, Y., Glocker, B., Davatzikos, C., Paragios, N.: Simultaneous Geometric - Iconic Registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 676–683. Springer, Heidelberg (2010)
13. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)

15. Yeo, B., et al.: Dt-refind: Diffusion tensor registration with exact finite-strain differential. IEEE Trans. Med. Imaging 28(12), 1914–1928 (2009)
16. Yeo, B., et al.: Spherical demons: Fast diffeomorphic landmark-free surface registration. IEEE Trans. Med. Imaging 29(3), 650–668 (2010)
17. Zhang, H., et al.: Deformable registration of diffusion tensor mr images with explicit orientation optimization. Medical Image Analysis 10(5), 764–785 (2006)

Automated Skeleton Based Multi-modal Deformable Registration of Head&Neck Datasets

Sebastian Steger and Stefan Wesarg

Cognitive Computing & Medical Imaging, Fraunhofer IGD, Darmstadt, Germany
{sebastian.steger,stefan.wesarg}@igd.fraunhofer.de

Abstract. This paper presents a novel skeleton based method for the registration of head&neck datasets. Unlike existing approaches it is fully automated, spatial relation of the bones is considered during their registration and only one of the images must be a CT scan. An articulated atlas is used to jointly obtain a segmentation of the skull, the mandible and the vertebrae C1-Th2 from the CT image. These bones are then successively rigidly registered with the moving image, beginning at the skull, resulting in a rigid transformation for each of the bones. Linear combinations of those transformations describe the deformation in the soft tissue. The weights for the transformations are given by the solution of the Laplace equation. Optionally, the skin surface can be incorporated. The approach is evaluated on 20 CT/MRI pairs of head&neck datasets acquired in clinical routine. Visual inspection shows that the segmentation of the bones was successful in all cases and their successive alignment was successful in 19 cases. Based on manual segmentations of lymph nodes in both modalities, the registration accuracy in the soft tissue was assessed. The mean target registration error of the lymph node centroids was 5.33 ± 2.44 mm when the registration was solely based on the deformation of the skeleton and 5.00 ± 2.38 mm when the skin surface was additionally considered. The method's capture range is sufficient to cope with strongly deformed images and it can be modified to support other parts of the body. The overall registration process typically takes less than 2 minutes.

Keywords: Image Registration, Head&Neck, Multi-Modal, Multi-Rigid.

1 Introduction

The registration of intra subject head&neck 3D datasets is required in many clinical applications. In image guided radiation therapy it enables adapting a previously generated plan to the patient's pose during intervention. Furthermore, it can be used for image fusion of different imaging modalities like Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). Thus, the clinical target volume and organs at risk can be delineated in the more suitable modality and propagated to the other modality, thereby improving the treatment plan. If spatial correspondence by the means of image registration is available, such a plan can also be propagated to a follow up CT scan which facilitates the plan

adaption. Beyond that, computer aided diagnosis systems can benefit from such an image fusion as they can incorporate different modalities for image feature extraction from clinically relevant targets such as tumors or lymph nodes.

Whereas rigid registration works well for the head, it is not able to cope with the neck due to possibly different positions of the spine. Intensity based deformable registration techniques typically have a smaller capture range and may produce poor results in regions with low contrast/signal or heavy imaging artifacts.

A popular approach to overcome these limitations is to exploit that the skeleton is the supporting structure of the soft tissue. The bones are treated as individual rigid bodies and deformations are allowed in the soft tissue only. Originally proposed by Little et al. [9], many other works are based on this principle. It is mostly applied to the registration of the spine [7][6][15][14], but also to the head&neck [5][1], the brain[3], the hand [10] and even to the whole body [8]. On Micro-CTs it is used for the registration of the lower limbs [11] and the whole body [4][13] of mice.

One of the key challenges of this principle is the identification of the individual rigid parts, i.e. the bones. Whereas most methods [1][3][5][9][11][15] depend on manual or interactive segmentation, [6][7] pursue semi-automatic approaches while [10][14] present very specific automatic heuristics. In [4], an articulated atlas is used. Methods not requiring explicit segmentation [8][13] of the individual bones can only be used if both images are CT scans from which the bone surface can easily be extracted.

During the registration of the individual bones the articulation is only considered in a few cases [4][5][10][11]. The deformation of the surrounding soft tissue can be computed solely based on the rigid transformations using thin plate splines (TPS) [7], linear combinations of transformations [9][14][10], a Log-Euclidean framework [3] or finite element methods (FEM) [5][1]. Alternatively, the rigid parts can be incorporated as constraints [15] or serve as an initialization [13][8] for an overall deformable intensity based registration. Rarely, the evaluation of the registration accuracy is carried out quantitatively. In [1], the center of volume error and dice similarities are reported for manually delineated targets whereas the distance of 20 landmarks is reported in [5]. Both evaluations are based on only 4 pairs of images.

In this paper, we present a fully automated registration method based on multi-rigid registration of the skeleton for 3D images of the head&neck. One of the images must be a CT image – the *fixed image* – but the other one – the *moving image* – can be from a different modality for which rigid registration methods exist or an image of the same subject acquired at a different point in time. After an overall mutual information based rigid registration of the head, the articulated atlas presented in [12] is used to obtain a segmentation of the individual bones from the CT image. These bones are then successively rigidly registered with the moving image resulting in a rigid transformation for each of them. Linear combinations of those transformations describe the soft tissue deformation. Optionally, the skin surface can be incorporated. The approach is quantitatively evaluated on 20 CT/MRI pairs of clinically acquired head&neck datasets.
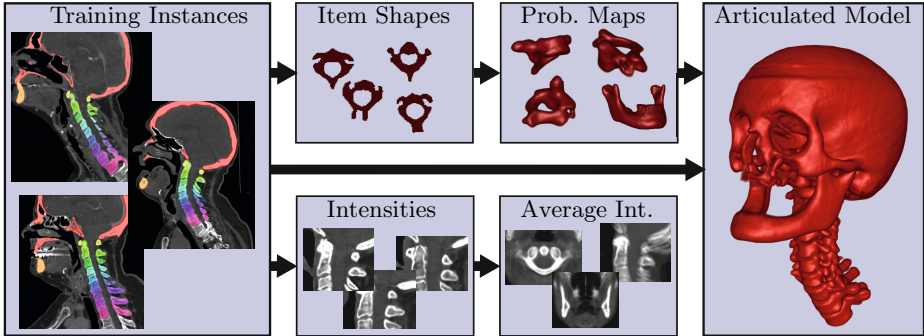
Fig. 1. Training of the articulated atlas for segmentation of bones from the fixed image

2   Methods

Articulated Atlas for CT Bone Segmentation. Due to its high contrast, bone can easily be segmented in CT images. However, since different bones have a similar appearance and are close to each other (e.g. the vertebrae), simple image segmentation techniques like thresholding are not able to distinguish neighboring bones from each other. Therefore, we use the articulated atlas presented in [12]. It is capable of jointly segmenting the skull, mandible, the cervical vertebrae and the two upper thoracic vertebrae. The atlas was created from head&neck CT datasets of 15 patients for which all bones had been labeled manually. Assuming bilateral symmetry of the bones, their right/left flipped versions were also used, resulting in a total of 30 training instances.

Unlike Baiker et al.'s articulated atlas [4], the articulation of the rigid parts is not explicitly modeled but learned from the training images. Therefore, particular attention was paid to cover a wide range of possible poses of the spine when selecting the training images.

For each type of bone, the probability map indicating the membership to the item along with the average intensities was computed in a coordinate system normalized by translation, rotation and scaling. Each bone item has those 7 degrees of freedom. Based on Alexa's *Linear Combination of Transformations* [2] a 7-dimensional vector is used to uniquely describe the pose of an item with respect to the model's global translation and rotation as a linear combination of 7 basis transformations. For a total of m bone items, the overall pose of each training instance j is described by a $7m$-dimensional vector $\boldsymbol{x}_j$.

The space of all possible articulations is then described by applying Principal Component Analysis (PCA) on the training articulation $X = (\boldsymbol{x_1}, .., \boldsymbol{x_n})$. An arbitrary articulation $\boldsymbol{x}$ can be described with a $k \ll 7m$ dimensional vector $\boldsymbol{b}$ as $\boldsymbol{x} = \bar{\boldsymbol{x}} + A \cdot \boldsymbol{b} + \boldsymbol{r}$, where $\bar{\boldsymbol{x}}$ is the average articulation of all training instances x_j and A is a matrix consisting of the eigenvectors of the covariance matrix. The more likely an articulation $\boldsymbol{x}$ is, the smaller the magnitude of $\boldsymbol{b}$ and $\boldsymbol{r}$ will be.

The atlas is adapted to a CT image by minimizing an energy functional depending on the 6 dimensional vector $\boldsymbol{t}$ describing a global rigid transformation and the $7m$ dimensional articulation parameter vector $\boldsymbol{x}$:

$$E(\boldsymbol{x}, \boldsymbol{t}) = E_{\text{external}}(\boldsymbol{x}, \boldsymbol{t}) + \lambda E_{\text{internal}}(\boldsymbol{x}) \ . \tag{1}$$

The external energy E_{external} ensures gray value similarity between the test image and the trained intensities of the atlas items and that the atlas items are close to high CT intensities (i.e. bones). The internal energy E_{internal} ensures that the atlas is within or at least close to the trained articulation space. The empirically determined parameter λ balances the external and internal energy. A gradient descent optimizer first finds the global transformation parameters $\boldsymbol{t}$ and then the articulation and external parameters $\boldsymbol{x}, \boldsymbol{t}$ jointly. Once the articulated model converged, the segmentations S_j for each bone item are extracted based on the probability maps and the CT intensities. Please refer to [12] for details.

Successive Rigid Alignment of the Bones. Based on the segmentations S_j of the bones in the CT image, they are rigidly aligned to the moving image by maximizing mutual information with a gradient descent optimizer. In order to incorporate not only the rather homogeneous part inside the bone, but also the texture at its boundary, the fixed image samples are taken from a slightly (5 mm) dilated version of S_j during the metric computation process as suggested in [14]. The success of local optimization techniques finding the desired local maximum depends on an initialization close to that very maximum. For highly deformed head&neck images, the same vertebra can be far away in both images and thus an independent registration is not feasible. Therefore, we exploit the connectivity of the bones expressed in the order in which the items are aligned. We start with the skull and use its resulting rigid transformation to initialize the registration of the mandible and the topmost vertebra (C1). The result of the latter is then used to initialize the next vertebra (C2) and so on, resulting in a top-down strategy. At the end, m rigid transformations T_j are obtained, one for each bone item j. Effectively denoting an overall transformation as a concatenation of several transformations enlarges the capture range.

Soft Tissue Motion Coupling. Two main constraints are applied when computing the transformation field $T(\boldsymbol{x})$ for each point in the fixed image: i) The known transformations in the rigid parts shall be incorporated (i.e. $T(\boldsymbol{x}) = T_j \Leftrightarrow \boldsymbol{x} \in S_j$) and ii) the transformation field shall be continuous (i.e. $T(\boldsymbol{x}+\boldsymbol{\epsilon}) \approx T(\boldsymbol{x})$ for small $|\boldsymbol{\epsilon}|$). Note that not only the translational part but also the rotational part shall be continuous, which is advantageous as argued in [9].

We denote the transformation at an arbitrary location $\boldsymbol{x}$ as a linear combination of the known transformations T_j, again using the approach presented in [2]: $T(\boldsymbol{x}) = \sum_{j=1}^{m} \Phi_j(\boldsymbol{x}) T_j$. To achieve continuity, each component of the m-dimensional coefficient field shall satisfy Laplace's partial differential equation (PDE): $\Delta \Phi_j = 0$, where Δ is the Laplace operator. Dirichlet boundary conditions $\Phi_j(\boldsymbol{x}) = \{1 \Leftrightarrow \boldsymbol{x} \in \partial S_j; 0 \Leftrightarrow \boldsymbol{x} \in \partial S_{i \neq j}\}$ are used at the boundaries of the bone items ∂S_j enforcing our first constraint. In a first step, the
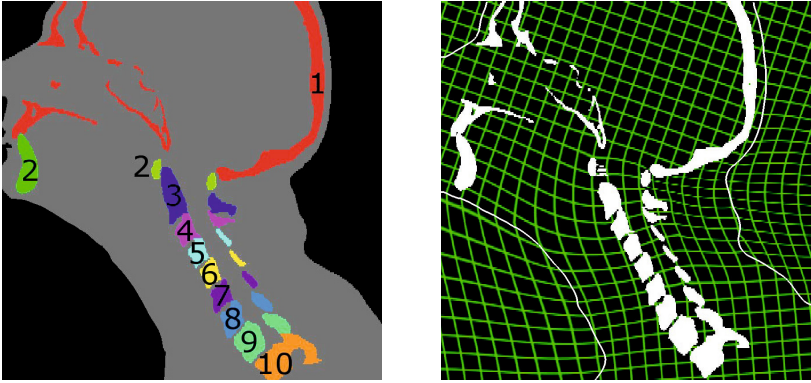
Fig. 2. Domain of the PDE in gray, registration order (left), deformed pattern (right)

domain Ω of the PDE is restricted to the body (gray region in Fig. 2) which is obtained by global thresholding. This ensures that transformations are only propagated within the body and undesired influences – such as from the chin directly into the thorax – are prevented. Here, Neumann boundary conditions are deployed ($\frac{\partial \Phi_j(\boldsymbol{x})}{\partial x_i} \mid_{\boldsymbol{x} \in \partial \Omega} = 0$). The solution of the PDE is found by solving the linear equation system resulting from finite differences approximation. In a second step, the domain is extended to the rest of the image (black region in Fig. 2) and previously obtained transformations are used as boundary condition.

Further Refinement. Since the computation of the soft tissue deformation solely depends on the rigid transformations of the bones, the continuity assumption may not result in the desired accuracy, especially if large deformations are present. One can imagine many different solutions to overcome this issue, such as a potentially more realistic deformation model (e.g. finite element methods [5][1]) or the incorporation of soft tissue image intensities into the registration process with rigidity constraints for the bones [15]. However, the approach we pursue is including further landmarks – the body surface – as additional constraints. Extracting the body surface in both images is done by simple thresholding. Since only small differences between both surfaces are expected, correspondences are found along the normal vectors of the fixed image's surface which are then used to correct the translational part of the deformation field in the entire soft tissue.

3   Experiments and Results

The presented fully automated registration method was evaluated on 20 intra-subject pairs of CT/MRI head&neck images of oral cancer patients acquired in clinical routine. The spacing between CT slices was $1 - 2$ mm. The MRI images consist of T1 weighted, fat saturated sagittal slices with a slice gap of $3 - 5$ mm. Some of the images showed heavy imaging artifacts.
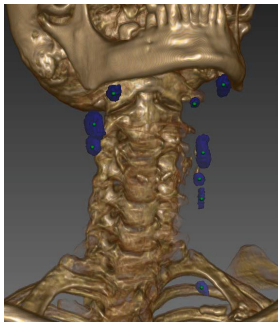
Fig. 3. Evaluation targets

Table 1. The minimum, maximum and average target registration error of lymph node centroids in mm

TRE	Rigid	Bone	Bone+Skin
Avg.	7.69 ± 4.05	5.33 ± 2.44	5.00 ± 2.38
Min.	1.33	0.37	0.92
Max.	61.21	26.20	25.32

At first, the segmentation of the bones in the CT images was assessed. Since the segmentation accuracy is not critical for the subsequent steps, visual inspection is sufficient to assess if all articulated atlas items ended up at the correct target structures. This was the case for all 20 CT-Images, indicating the robustness of the deployed articulated atlas. The successive rigid alignment was assessed in the same way. Except for one single MRI image, all rigid structures were aligned correctly. The reason for the failure in that one case was a very low MRI signal located directly in the spine. Thus, one vertebra converged to the wrong local minimum, resulting in a design based misalignment of all subsequently aligned vertebrae.

Then, in a second experiment, the soft tissue accuracy was assessed for the 19 successful aligned image pairs. Since the evaluation was based on clinical images, neither a ground truth deformation field nor artificial corresponding landmarks were available. Therefore, we chose to measure the target registration error (TRE). As target landmarks we used the centroid of lymph nodes, which were manually delineated in both imaging modalities. On the average we found 5.2 clearly corresponding targets per dataset. The reason we chose lymph nodes is that they are clinically relevant, reasonably well distributed in the soft tissue of the head&neck region (see Fig. 3), locally bound and visible in both imaging modalities. Due to the images' different resolutions and the resulting segmentation differences the centroid appeared to be more robust than e.g. the average surface distance. The average and maximum TRE was separately computed for the bone-aligned and the surface-refined images as well as for the rigidly head-aligned image for comparison. The results are presented in Table 1 and an example for a registration result for a strongly deformed pair of images can be seen in Fig. 4. Whereas targets close to the bones resulted in high accuracies by design, the worst accuracy were achieved very far away from the head. There, different positions of the shoulder – not part of the skeleton model – had an impact, explaining the high maximum TREs. Further inaccuracies can be attributed to inaccurate segmentations due to the low inter slice resolution of the MRI images. Since only primitive image processing techniques were deployed to

(a) rigid head alignment   (b) bone only   (c) bone and surface
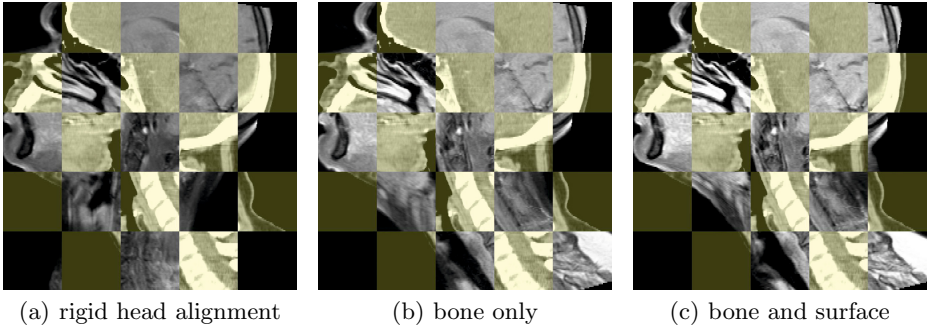
Fig. 4. CT/MRI Registration of a strongly deformed neck (fixed CT colored)

detect the skin surface, this step was locally not successful for some MRI images, resulting in geometric distortion. Model based approaches may be required.

To give a rough indication about the runtime performance of our method, we measured the duration of our prototypical implementation for each individual step on a Intel Core i7-2600K CPU, 3.4Ghz machine with 16GB memory. On the average it took 107s to finish the overall registration, where the rigid registration accounted for 14%, the bone segmentation for 32%, the successive alignment for 16%, the soft tissue motion coupling for 22% and the refinement step for 16%. We are convinced that a considerable speedup is achievable with an optimized implementation exploiting parallelization.

4   Conclusion

In this paper we presented a skeleton based deformable registration approach and applied it to head&neck CT/MRI datasets. Unlike existing approaches, it is fully automated, supports all modalities for which rigid registration techniques exist, is able to cope with strong deformations and was evaluated quantitatively on images acquired in clinical routine. The mean target registration error of the lymph node centroids was 5.33 ± 2.44 mm when the registration was solely based on the deformation of the skeleton and 5.00 ± 2.38 mm when the skin surface was additionally considered. Whereas the articulated atlas enables the automation, the support of strongly deformed image pairs is ensured by the concatenation of several transformations to a single transformation of greater magnitude during the successive alignment of the individual bone structures. With a runtime of less than 2 minutes it is suitable for the use in clinical routine.

Rebuilding the articulated atlas, the approach can be extended to other parts of the body (e.g. the thorax or the pelvis). Without the refinement step, the presented approach has another interesting property when applied to follow up CT scans. Under the assumption that the bones remain constant over time, the method does to some extent compensate for deformations induced by a change in

position only, but not for changes in anatomy (e.g. tumor growth or surgery) because the soft tissue image intensities are not considered. Whereas rigid registration has this property as well, almost no deformable intensity based registration approaches are able to distinguish between those two sources of deformation at all. This may be disadvantageous in some applications (e.g. propagating a radiation therapy plan), but it enables detecting differences in follow up images.

Future work includes increasing the robustness for images with imaging artifacts, deploying more realistic soft tissue motion coupling, applying the presented scheme to other body parts and the automated detection of differences.

References

1. Al-Mayah, A., et al.: Biomechanical-based image registration for head and neck radiation treatment. Phys. Med. Biol. 55(21), 6491 (2010)
2. Alexa, M.: Linear combination of transformations. In: SIGGRAPH 2002, pp. 380–387. ACM, New York (2002)
3. Arsigny, V., et al.: A fast and log-euclidean polyaffine framework for locally linear registration. J. Math Imaging Vis. 33, 222–238 (2009)
4. Baiker, M., et al.: Fully automated whole-body registration in mice using an atriculated skeleton atlas. In: Proc IEEE Int. Symp. Biomed. Imaging, pp. 728–731 (2007)
5. du Bois dAische, A., et al.: Estimation of the deformations induced by articulated bodies: Registration of the spinal column. Biomed Signal Proces. 2(1), 16–24 (2007)
6. Hu, Y., Haynor, D.R.: Multirigid registration of mr and ct images of the cervical spine. In: Fitzpatrick, J.M., Sonka, M. (eds.) Medical Imaging 2004: Image Processing, vol. 5370, pp. 1527–1538. SPIE (2004)
7. Huesman, R., et al.: Deformable registration of multi-modal data including rigid structures. In: IEEE Nucl. Sci. Symp. Conf. Rec., vol. 3, pp. 1879–1882 (2002)
8. Li, X., Peterson, T.E., Gore, J.C., Dawant, B.M.: Automatic Inter-subject Registration of Whole Body Images. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 18–25. Springer, Heidelberg (2006)
9. Little, J., et al.: Deformations incorporating rigid structures. In: Proc. Math. Methods in Biomed. Image Anal., pp. 104–113 (June 1996)
10. Martín-Fernández, M.A., et al.: Automatic articulated registration of hand radiographs. Image Vision Comput. 27(8), 1207–1222 (2009)
11. Papademetris, X., Dione, D.P., Dobrucki, L.W., Staib, L.H., Sinusas, A.J.: Articulated Rigid Registration for Serial Lower-Limb Mouse Imaging. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 919–926. Springer, Heidelberg (2005)
12. Steger, S., et al.: Articulated atlas for segmentation of the skeleton from head & neck ct datasets. In: Proc IEEE Int. Symp. Biomed. Imaging, pp. 1256–1259 (2012)
13. Suh, J.W., et al.: A non-rigid registration method for serial lower extremity hybrid spect/ct imaging. Med. Image Anal. 15(1), 96–111 (2011)
14. Čech, P., et al.: Piecewise rigid multimodal spine registration. In: Handels, H., Ehrhardt, J., Horsch, A., Meinzer, H.-P., Tolxdorff, T. (eds.) Bildverarbeitung für die Medizin 2006, pp. 211–215. Informatik aktuell, Springer, Heidelberg (2006)
15. Wang, K., He, Y., Qin, H.: Incorporating Rigid Structures in Non-rigid Registration Using Triangular B-Splines. In: Paragios, N., Faugeras, O., Chan, T., Schnörr, C. (eds.) VLSM 2005. LNCS, vol. 3752, pp. 235–246. Springer, Heidelberg (2005)

Lung Registration with Improved Fissure Alignment by Integration of Pulmonary Lobe Segmentation

Alexander Schmidt-Richberg, Jan Ehrhardt, René Werner, and Heinz Handels

Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
`schmidt-richberg@imi.uni-luebeck.de`

Abstract. Accurate registration of human lungs in CT images is required for many applications in pulmonary image analysis and used for example for atlas generation. While various registration approaches have been developed in the past, the correct alignment of the interlobular fissures is still challenging for many reasons, especially for inter-patient registration. Fissures are depicted with very low contrast and their proximity in the image shows little detail due to the lack of vessels. Moreover, iterative registration algorithms usually require the objects to be overlapping in both images to find the right transformation, which is often not the case for fissures.

In this work, a novel approach is presented for integrated lobe segmentation and intensity-based registration aiming for a better alignment of the interlobular fissures. To this end, level sets with a shape-based fissure attraction term are used to formulate a new condition in the registration framework. The method is tested for pairwise registration of lung CT scans of nine different subjects and the results show a significantly improved matching of the pulmonary lobes after registration.

1 Introduction

Establishing correspondences between two thoracic CT images by intensity-based non-linear registration is a frequent task in various clinical applications. It is required for example for estimating the pulmonary lung motion of patients using 4D image data or to align two scans of one patient in follow-up studies. Moreover, atlas generation techniques are generally based on registration algorithms to match scans of different patients.

Various approaches for lung registration have been proposed in the past. To compare their performance, extensive evaluation studies have recently been conducted [1,2]. While they demonstrate a generally high accuracy of the different algorithms, results also indicate that the interlobular lung fissures are often not sufficiently aligned. Murphy et al. state in their study [2] that while "most of the algorithms performed extremely well in terms of both singularities and lung boundary alignment, [...] differences are much more apparent in the fissure alignment category." This has mainly two reasons: First, the fissures are often depicted with very low contrast – especially in low-dose CT images – and

intensity-based registration is therefore difficult. There are also few vessels in the proximity of fissures, which leads to homogeneous greyvalues in this region. Secondly, to steer iterative algorithms in the right direction, structures usually have to be overlapping. While multi-scale approaches and regularization schemes implicitly address these problems, small structures such as the fissures are not visible on coarse scales and no forces are explicitly generated that align disjunct structures as caused by large displacements or anatomical dissimilarities. This is especially severe for inter-patient registration.

Several recent publications propose combining image registration and a segmentation of the object-of-interest to improve registration in critical regions [3,4]. However, these methods use solely intensity-based segmentation methods that are not applicable for lobe segmentation. A registration of distance maps generated from lung boundaries and fissure segmentations was therefore proposed in [5], but image intensities are not considered.

In this work, we present an approach for including explicit fissure alignment in intensity-based image registration algorithms. Assuming a known lobe segmentation of one image (e.g. an atlas), automatic segmentation of the second image and registration are incorporated in a joint approach. Their mutual dependency is used to combine intensity- as well as shape information to guide the registration. An integrated variational formulation of the problem is presented. The approach is preliminarily evaluated for pairwise inter-patient registration using nine thoracic CT images.

2   Methods

We proceed by briefly introducing an intensity-based non-linear registration scheme (Section 2.1). An approach for automatic lobe segmentation based on multi-object level sets was proposed in [6] and is summarized in Section 2.2. In Section 2.3, these methods are then integrated in a joint framework.

2.1   Variational Lung Registration

Given two images I^R and I^T (called reference and template image) with $I : \Omega \mapsto \mathbb{R}$ and the image domain $\Omega \subset \mathbb{R}^3$, the registration problem can be formulated as finding a plausible transformation $\varphi : \Omega \mapsto \Omega$ that transforms the template image to match the reference image by minimizing the energy functional

$$\mathcal{J}^{Reg}[\varphi] := \mathcal{D}[I^R, I^T; \varphi] + \alpha_1 \mathcal{S}[\varphi] . \tag{1}$$

Here, $\mathcal{D}$ is a distance measure quantifying the (dis-)similarity between reference and transformed target image. The plausibility of the transformation is controlled by the regularizer $\mathcal{S}$, which smoothes the field and thereby avoids discontinuities like gaps or folding, and α_1 weights the amount of smoothing.

With regard to the Euler-Lagrange equation of the functional, the transformation φ can be optimized using a gradient descent according to

$$\frac{\partial \varphi}{\partial t} = \boldsymbol{f}(\boldsymbol{u}) + \alpha_1 \mathcal{A}\boldsymbol{u} , \tag{2}$$

where $\boldsymbol{u}(\boldsymbol{x}) := \varphi(\boldsymbol{x}) - \boldsymbol{x}$ denotes the displacement field of the transformation, $\mathcal{A}$ is a linear operator related to the regularizer and $\boldsymbol{f}$ a force term derived from the distance measure. Without loss of generality, diffusion regularization and normalized SSD-forces (NSSD) are used in this work [7].

2.2   Lobe Segmentation with Multi-object Level Sets

In level set segmentation, the boundary of an object $\Sigma \subset \Omega$ in the image is represented implicitly as the zero-level curve of a level set function $\phi : \Omega \mapsto \mathbb{R}$. Here, ϕ is defined as the distance function to the boundary with $\phi(\boldsymbol{x}) < 0$, if $\boldsymbol{x} \in \Sigma$ and $\phi(\boldsymbol{x}) > 0$, if $\boldsymbol{x} \in \Omega \setminus \Sigma$. The optimal level set is determined by minimizing the energy functional

$$\mathcal{J}^{Seg}[\phi] := \mathcal{E}[I; \phi] + \alpha_2 \mathcal{I}[\phi] . \tag{3}$$

The internal energy $\mathcal{I}$ is defined as in [6] and enforces a smooth surface. The region-based external energy

$$\mathcal{E}[I; \phi] := - \int_\Omega (1 - H(\phi(\boldsymbol{x}))) \log p_{in}(I(\boldsymbol{x})) \; + \; H(\phi(\boldsymbol{x})) \log p_{out}(I(\boldsymbol{x})) \; d\boldsymbol{x}$$

integrates a-priori knowledge about intensity distributions p_{in} inside and p_{out} outside the lungs, respectively, and draws the segmentation to the lung boundaries. The Heaviside function H is used to distinguish inside from outside.

For the minimization with respect to the level set function, the Euler-Lagrange equation is derived and a gradient descent is performed according to

$$\frac{\partial \phi}{\partial t} = -\delta(\phi) \left(\log \frac{p_{in}}{p_{out}} - \alpha_2 \nabla \frac{\nabla \phi}{\|\nabla \phi\|} \right) .$$

To simultaneously segment the pulmonary lobes, the level set framework is extended following Brox et al. [8] by employing N functions $\phi_i, i = 0, \ldots, N-1$, each representing one object $\Sigma_i := \{\boldsymbol{x} : \phi_i(\boldsymbol{x}) < 0\}$. Front propagation is then performed according to

$$\frac{\partial \phi_i}{\partial t} = -\delta(\phi) \left(e_i - \max_{H(\phi_j)<0, j \neq i} (e_j, e_i - 1) \right) \text{ with } e_k := \log p_k - \frac{\alpha_2}{2} \nabla \frac{\nabla \phi_k}{\|\nabla \phi_k\|} . \tag{4}$$

Here, e_i serves as a (mostly outwards-directed) force that affects the function ϕ_i. The final update value is then determined by a competition of this force and the maximal force of all adhering level set functions ϕ_j. For lobe segmentation, we have $N = 6$ objects (background and five lobes) and set $p_0 := p_{out}$ and $p_i := p_{in}$ for $i = 1, \ldots, 5$.

Using the model described above, only smoothing is performed between two lobes because p_i equals p_j in these cases. Therefore, we follow [6] and define an additional term that draws the contour to the fissures. With this purpose, the interlobular fissures are segmented employing the supervised enhancement filter
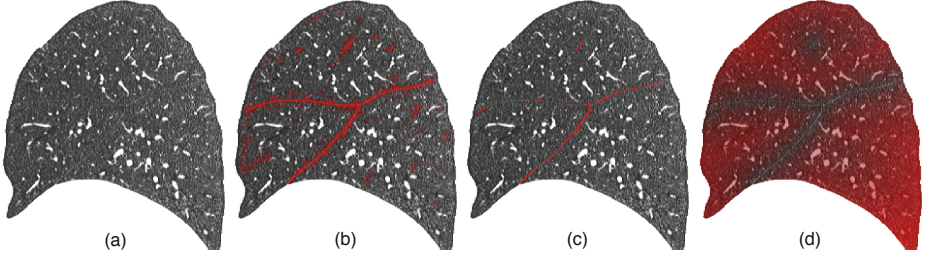
Fig. 1. Computation of the cost image for lobe segmentation: (a) the CT image I; (b) the fissure segmentation as computed by the supervised enhancement filter; (c) the skeleton of the post-processed fissure segmentation (dilated by one voxel to enhance visibility); (d) the cost image C, which is incorporated in the force term

proposed by van Rikxoort et al. [9], which uses intensity and shape information to train a kNN classification for the detection of fissure voxels. To reduce false positives, parameters are chosen aiming at a high specificity at the expense of sensitivity. A cost image $C : \Omega \mapsto \mathbb{R}$ is computed as distance transformation to the skeleton K of the fissure segmentation (see Figure 1). With this, the force term in (4) can be reformulated to incorporate a fissure-attraction force by

$$e_k := \log p_k - \frac{\alpha_2}{2} \nabla \frac{\nabla \phi_k}{\|\nabla \phi_k\|} - \frac{\beta}{2} \nabla \phi_k \cdot \nabla C \ . \tag{5}$$

2.3 Integrated Lobe Segmentation and Registration

Registration and lobe segmentation are combined following the principle idea of [4] and extending it by multi-object lobe segmentation. In the procedure, segmentations ϕ_i^T of the lobes in the template image are assumed to be known. The approach aims at simultaneously finding a transformation φ and a segmentation ϕ_i^R of the reference image under the additional condition that the transformed template segmentation resembles the segmentation of the reference image.

The integrated model is defined by the joint energy functional

$$\mathcal{J}^{Joint}[\varphi, \phi^R] := \mathcal{J}^{Reg}[\varphi] + \gamma \mathcal{J}^{Seg}[\phi^R] + \gamma_1 \mathcal{P}[\phi^T; \varphi, \phi^R] \ , \tag{6}$$

where the segmentation- and registration-related terms are defined as before and γ and γ_1 weight the terms against each other. The shape prior term

$$\mathcal{P}[\phi^T; \varphi, \phi^R] := \frac{1}{2} \int_\Omega \sum_{i=0}^{N-1} \delta(\phi_i^R) \Big(\phi_i^T \circ \varphi - \phi_i^R \Big)^2 \ d\boldsymbol{x}$$

links segmentation and registration by penalizing large distances between the zero level set of ϕ^R and the transformed template segmentation $\phi_i^T \circ \varphi$. It is formulated exploiting the fact that the value of a level set function is defined to be the distance to the closest boundary.
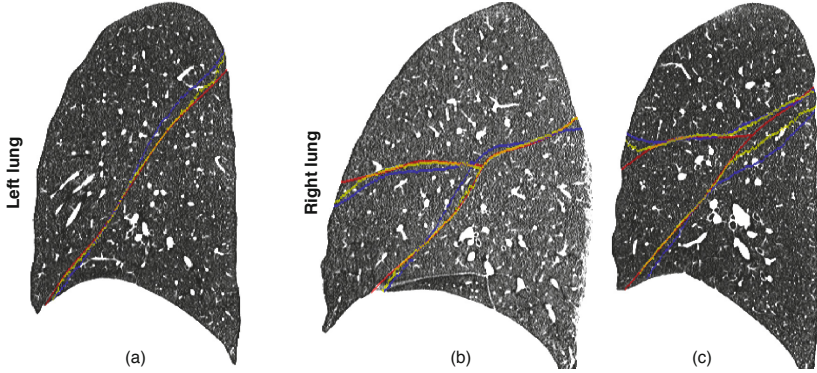
Fig. 2. Reference images with lobe boundaries as color overlay. Red: manual reference segmentation; Blue: template segmentation after registration with $\mathcal{J}^{Reg}$; Yellow: template segmentation after registration with $\mathcal{J}^{Joint}$. In the left lung (a), alignment is improved for almost all patients. For the right lung, results depend on the initialization: While fissures are well-aligned in (b), alignment fails with both algorithms in (c).

The joint energy term is alternately minimized in direction of the transformation and of the reference segmentation. Including the shape prior term, the gradient descent equation (2) for the registration part reads

$$\frac{\partial \varphi}{\partial t} = \boldsymbol{f}(\boldsymbol{u}) + \alpha_1 \mathcal{A}\boldsymbol{u} + \gamma_1 \sum_{i=0}^{N-1} \delta(\phi_i^R)\Big(\phi_i^T \circ \varphi - \phi_i^R\Big)\nabla(\phi_i^T \circ \varphi) \ ,$$

where the last term corresponds to the Euler-Lagrange equation of $\mathcal{P}$. It acts as an additional force that affects φ to transform the template segmentation in the direction of the reference segmentation. The segmentation is updated according to (4) but using the force

$$e_k := \log p_k - \frac{\alpha_2}{2}\nabla\frac{\nabla\phi_k^R}{\|\nabla\phi_k^R\|} - \frac{\beta}{2}\nabla\phi_k^R \cdot \nabla C + \frac{\gamma_2}{2}(\phi_k^T \circ \varphi - \phi_k^R)$$

with $\gamma_2 := \gamma_1/\gamma$ for force computation. Here, the additional term prevents the segmentation from diverging too far from the transformed template segmentation and thus avoids leakage into neighboring structures.

3   Results

The approach is evaluated for inter-patient registration on a set of nine thoracic normal dose CT images of different subjects (120 kVp, 450-750 mAs, 0.79×0.79× 0.7 mm spacing, cropped to lung region). For each image, lobe segmentations S where generated for evaluation using a spline-based interpolation of manually defined fissure points and a subsequent manual correction. To avoid a bias due

Table 1. Lobe and fissure alignment after registration with common intensity-based registration $\mathcal{J}^{Reg}$ and the proposed registration with fissure alignment $\mathcal{J}^{Joint}$, averaged over the results of all 72 image pairs. The mean fissure distance is given in millimeter.

Lobe	Dice coeff. $\mathcal{J}^{Reg}$	$\mathcal{J}^{Joint}$	Fissure	Metric $Fiss$ $\mathcal{J}^{Reg}$	$\mathcal{J}^{Joint}$	Mean dist. $\mathcal{J}^{Reg}$	$\mathcal{J}^{Joint}$
left superior	0.92	0.95	left	0.18	0.52	9.79	4.74
left inferior	0.93	0.95					
right superior	0.88	0.90	right oblique	0.16	0.44	11.34	6.10
right middle	0.69	0.73	right horizontal	0.15	0.24	14.72	12.70
right inferior	0.91	0.94					

to the choice of the reference image, a pairwise validation is employed, that means each image is registered with the remaining eight images leading to 72 registrations. Parameters of the segmentation and registration components were chosen as specified in [6,7]. The values for the additional parameters γ_1 and γ_2 were optimized empirically. Since the segmentation converges faster than the registration, five registration iterations are performed for each segmentation iteration (this parameter can be seen as related to γ in (6)). In total, computation time lies between 6 min and 14 min, which means a prolongation of approximately 48% in comparison to the standard registration.

For a quantitative evaluation, the Dice coefficient of the manual reference segmentation S^R and transformed template segmentation $S^T \circ \varphi$ were calculated after registration with the common approach using $\mathcal{J}^{Reg}$ and the presented registration with integrated lobe segmentation $\mathcal{J}^{Joint}$ (see Table 1). Since the Dice coefficient may be difficult to interpret due to the varying size of the individual lobes, two additional metrics that quantify fissure alignment are considered: on the one hand, the mean distance of fissure voxels in $S^T \circ \varphi$ to fissure voxels S^R; on the other hand, a metric inspired by Murphy et al. [2] that estimates the proportion of the lobe boundaries in the reference image that lie in the proximity of the lobes in the transformed template image. For this, let $B(S^T \circ \varphi)$ be the set of voxels in the transformed template segmentation S^T that lie at the boundary between two lobes. To incorporate some tolerance, this region is expanded by ± 3 voxels in z-direction and denoted by B^+. The proportion of correctly segmented boundary voxel X can then be quantified by

$$Fiss(S^T \circ \varphi, S^R) := \frac{|\{X : X \in B^+(S^T \circ \varphi) \wedge X \in B(S^R)\}|}{|\{X : X \in B(S^R)\}|} \ .$$

The results confirm a significantly better alignment of the lung fissures using the proposed registration approach (paired t-test, $p \leq 0.05$ for all metrics and subjects serving as reference image). Averaged over all registrations, 40% of the fissures are well aligned using the joint approach, compared to only 16% with the standard method. Results are better for the left lung (52%) than for the fissures in the right lung (44% and 24% for horizontal and oblique fissures). The same observations hold for the mean fissure distance.
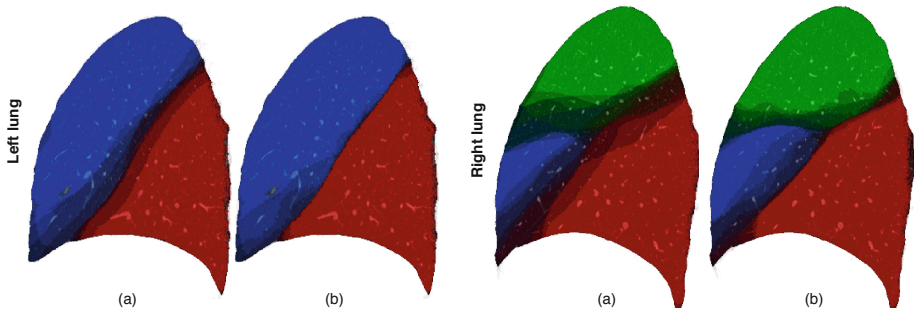
Fig. 3. Transformed lobe segmentations of eight subjects dealing as atlas after registration with the remaining subject. The joint approach (b) produces a sharper and therefore more consistent atlas segmentation than the standard registration (a).

The results are visualized in Figure 2. A much better alignment of the fissures is observed in regions where fissures were segmented successfully. Furthermore, in a proof-of-concept demonstration the application to atlas generation is explored and illustrated in Figure 3. Here, lobe segmentations of eight subjects were transformed to the remaining subjects and summed up. The joint registration approach provides a much sharper segmentation of the reference image, which indicates a superior matching of the lobes.

4   Discussion

The results show that integrating lobe information in intensity-based registration can improve fissure alignment considerably. Employing level set methods for this entails several advantages over – for example – directly considering the cost image C. First, forces are only generated along the zero level set, that means only at the fissures and not at places wrongly classified as fissures due to noise. Gaps in the fissure segmentations are not exceedingly critical since they are automatically bridged by the level set framework. If no fissure information is present in an image region, no forces (beside smoothing) are generated that move ϕ^R away from ϕ^T and thus only the standard registration is applied in this area. This effect could be enhanced by truncating C to restrain the influence of the fissure alignment to a smaller range, for example if segmentation is suffering from bad image quality.

Problems arise in particular in the right lung between horizontal and oblique fissures because the level set can be attracted by the wrong fissure if the initialization is bad. Moreover, the fissure enhancement filter performs worse in this region. A more precise initialization – for example using an anatomical atlas [6] – could improve the results in this area.

Another interesting application of the approach would be the estimation of breathing motion in 4D image data. While the displacement is smaller in this case, fissure segmentation is very difficult in 4D images due to the often lower

radiation dose per image as well as motion and reconstruction artifacts. These aspects will be the focus of future investigations.

5   Conclusion

In this work, an approach is presented to include pulmonary lobe segmentation in intensity-based registration with the aim of improving the alignment of the interlobular fissures. Segmentation is done using multi-object level sets with an additional shape-based fissure attraction term. This information is used in a new condition in the registration framework to match the lobes to each other.

The approach was tested for inter-patient registration on nine thoracic CT scans. A considerably better alignment of the fissures was observed but the method depends on a reasonably good fissure segmentation. Moreover, huge anatomical differences between the patients imply a bad initialization of the segmentation, which may result in a misalignment of the fissures. This is especially critical for the right lung, where horizontal and oblique fissures are close to each other.

References

1. Brock, K.K.: Deformable Registration Accuracy Consortium: Results of a multi-institution deformable registration accuracy study (MIDRAS). Int. J. Radiat. Oncol. Biol. Phys. 76(2), 583–596 (2010)
2. Murphy, K., van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., et al.: Evaluation of Registration Methods on Thoracic CT: The EMPIRE10 Challenge. IEEE Trans. Med. Imag. 30(11), 1901–1920 (2011)
3. Yezzi, A., Zöllei, L., Kapur, T.: A variational framework for integrating segmentation and registration through active contours. Med. Image Anal. 7(2), 171–185 (2003)
4. Schmidt-Richberg, A., Handels, H., Ehrhardt, J.: Integrated segmentation and non-linear registration for organ segmentation and motion field estimation in 4D CT data. Methods Inf. Med. 48(4), 344–349 (2009)
5. van Rikxoort, E.M., Prokop, M., de Hoop, B.J., Viergever, M.A., Pluim, J.P.W., van Ginneken, B.: Automatic Segmentation of Pulmonary Lobes Robust against Incomplete Fissures. IEEE Trans. Med. Imag. 29(6), 1286–1296 (2010)
6. Schmidt-Richberg, A., Ehrhardt, J., Wilms, M., Werner, R., Handels, H.: Pulmonary Lobe Segmentation with Level Sets. In: Haynor, D.R., Ourselin, S. (eds.) Proc. SPIE, p. 83142V (2012)
7. Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Handels, H.: Diffeomorphic Diffusion Registration of Lung CT Images. In: van Ginneken, B., Murphy, K., Heimann, T., Pekar, V., Deng, X. (eds.) Medical Image Analysis for the Clinic: A Grand Challenge, MICCAI 2010, pp. 55–62 (2010)
8. Brox, T., Weickert, J.: Level Set Segmentation with Multiple Regions. IEEE Trans. Image Process. 15(10), 3213–3218 (2006)
9. van Rikxoort, E.M., van Ginneken, B., Klik, M., Prokop, M.: Supervised Enhancement Filters: Application to Fissure Detection in Chest CT Scans. IEEE Trans. Med. Imag. 27(1), 1–10 (2008)

3D Ultrasound-CT Registration in Orthopaedic Trauma Using GMM Registration with Optimized Particle Simulation-Based Data Reduction

Ilker Hacihaliloglu[1], Anna Brounstein[2], Pierre Guy[1],
Antony Hodgson[3], and Rafeef Abugharbieh[2]

[1] Departments of Orthopaedics,
[2] Departments of Electrical and Computer Engineering,
[3] Mechanical Engineering,
University of British Columbia, Vancouver, BC, Canada
ilker.hacihaliloglu@hiphealth.ca, {abrou,rafeef}@ece.ubc.ca,
pierre.guy@ubc.ca, ahodgson@mech.ubc.ca

Abstract. Accurate real-time registration of intra-operative ultrasound (US) to computed tomography (CT) remains a challenging problem. In orthopedic applications, a recent promising approach proposed the use of Gaussian mixture modeling for bone surface registration. Though relatively successful, the method relied on naïve and error prone subsampling of the surfaces registered to reduce computational cost and also heavily relied on heuristically-set parameters for bone surface generation. In this paper, we present an improved approach employing a novel point simplification method that redistributes surface points to better represent the surface achieving near real-time registration with higher accuracy and robustness. We also present a framework for automating the parameter selection in the bone surface extraction step. For validation, we present extensive quantitative tests on phantom and clinical data obtained by scanning patients with pelvic ring fractures in the operating room. We show an 89% average improvement in target registration error over the recent GMM registration based method.

Keywords: intra-operative volume registration, 3D ultrasound to CT registration, orhopaedic imaging, real-time registration.

1 Introduction

Registering tracked intra-operative US images with pre-operative CT data has been proposed as a mechanism for making pre-operative CT more readily available for a range of computer assisted orthopaedic surgery (CAOS) procedures. The most widely used registration method in CAOS applications is the iterative closest point (ICP) algorithm, but since the initial publication of the ICP, new methods and algorithms have been proposed to improve the robustness and speed of the standard ICP algorithm [1]. Although relatively successful, ICP is susceptible to converging to local minima and therefore a close initial manual alignment is necessary. Moghari [2] proposed a point-based registration method based on the Unscented Kalman Filter

(UKF). Although successful registration results were achieved, with a mean registration error of 0.3mm, the registration time exceeded 20 seconds due to the higher computational complexity of the method. The bone surfaces were also extracted manually from US images before the registration process, which would preclude direct use of this technique intraoperatively. Penney et al. [3] used a normalized cross-correlation similarity metric to register bone probability images obtained from CT and US data sets. They reported a mean target registration error (TRE) of 2.3 mm for cadaver study where the registration time was between 2 and 10.5 min, which is well over the time needed for providing real-time guidance. Gill et al. [4] simulated US images from CT data for registering bone surfaces of the spine and achieved a registration accuracy of 1.44 mm for phantom scans and 1.25 mm for sheep cadaver scans. However their intensity-based registration took an average of 14 minutes on a central processing unit (CPU) and 11 seconds when implemented on a graphics processing unit (GPU). This approach was later extended to register statistical shape models (SSMs) of the lumbar spine where a TRE less than 3mm for phantom scans was reported [5]. Again, the registration time was on the order of hours. Recently, Brounstein et al. [6] proposed a Gaussian Mixture Model-based (GMM) surface registration algorithm. The bone surfaces were automatically extracted using local phase image features obtained by convolving the US volumes with 3D Log-Gabor filters where the filter parameters were selected empirically [7]. In order to improve the speed of the proposed algorithm, the extracted point clouds were reduced using a simple downsampling approach that kept only 5% of the surface points [6]. The proposed method was evaluated on a phantom setup and three clinical scans. Although the early results were promising, some significant problems remained. For example, while local phase-based techniques successfully extract the desired image features, they remain sensitive to the underlying filter parameters used. Furthermore, the registration could fail if key features were lost in the down sampling process.

In this work, we propose and test several improvements to this work, including (1) extracting bone surfaces using automatically optimized 3D Log-Gabor filter parameters, (2) using a novel point cloud simplification method that ensures fast run time and retains salient features needed to provide more robust and accurate registration results; and (3) introducing an optimization method that reduces the complexity of point simplification by 50%. We present validation studies not only on phantom data but also on an extensive set of clinical scans obtained from 21 human subjects with pelvic fractures to assess registration accuracy and robustness in the presence of typical US imaging artifacts.

2   Methods

2.1   US-CT Registration Using Gaussian Mixture Models (GMMs)

Gaussian Mixture Models (GMMs) are a way to replace a set of points in a multidimensional space with a (typically smaller) set of multi-dimensional Gaussian distributions (components) to describe sub-populations. Each component's density, φ_i, is characterized by its mean, μ_i, and its covariance matrix, Σ_i. By representing the

point sets as GMMs, we can compute an L2 distance metric between two different GMMs without the need to solve a computationally challenging point-to-point corres-pondence problem, which is normally required in many registration methods. The registration algorithm then minimizes this L2 distance between the model, $M(x) = p_m(x)$, and the scene, $S(x) = p_s(x)$ over the set of possible rigid transformations, $T(M(x), \theta, t)$, where θ is a 6-vector representing a rigid transform and the *L2* distance is given as:

$$d_{L2}(M(x), S(x), \theta, t) = \int \big(S(x) - T(M(x), \theta, t)\big)^2 dx. \tag{1}$$

It should be noted that the scene model, $S(x)$, is fixed during the optimization. Since $T(M(x), \theta, t)$ and $\int T(M(x), \theta, t)^2 dx$ are invariant for rigid transformations, mini-mizing the L2 distance given in Equation 1 becomes equivalent to solving:

$$argmin_{\theta, t}[d_{L2}(M(x), S(x), \theta, t)] = argmin_{\theta, t}\left[-\int S(x)T(M(x), \theta, t)dx\right] \tag{2}$$

where $\int \emptyset(x|\mu_1, \Sigma_1) \, \emptyset(x|\mu_2, \Sigma_2)dx = \emptyset(0|\mu_1 - \mu_2, \Sigma_1 + \Sigma_2)$. A closed form ex-pression for the L2 distance between GMMs can be found in [8]. In [6], this approach was used for registering CT scans to local phase bone surfaces extracted from US volumes using 3D Log-Gabor filters with empirical filter parameters. The CT vo-lumes were segmented using a binary threshold at 200 H.U [6]. Ray-casting was then used to find the bone surface in these binary CT volumes resulting in extracted sur-faces that are one voxel thick. In order to achieve real-time registration the extracted bone surfaces were reduced using a simple/naïve down sampling approach by keeping only 5% of the total surface points. Finally, Gaussian curvature information was incorporated back into the reduced point clouds to provide correct registration along high curvature bone surface areas [6]. A critical consideration for successful local phase based feature extraction in medical images is the proper configuration of the different and highly sensitive parameters involved. Parameter values typically need to be scene-adaptive in order to accurately localize structures of interest in the scanned areas. This is especially important for extracting bone surfaces in US since in clinical scenarios the response of soft tissue interfaces can often resemble that of the bone. Furthermore, naïve down sampling can be problematic as it may not capture all the anatomical features the bone surface can offer, nor provide an even spread in the point cloud. Furthermore, it will allow outliers to influence the shape of the point cloud more and it is even possible for entire planes of the bone surface to be omitted, requir-ing more points in each point cloud to represent the surface which dramatically in-creases the runtime. In the next sections we provide solutions to these problems by proposing a framework for automatic point cloud optimization.

2.2   Bone Surface Extraction Using Optimized 3D Log-Gabor Filters

The use of 3D phase symmetry (3DPS) to automatically extract bone surfaces in US based on Log-Gabor filters was originally proposed in [7]. The transfer function of a 3D Log-Gabor filter, *3DG_{ij}*, in the frequency domain can be constructed as the prod-uct of two components (3); a one dimensional Log Gabor function controlling the

frequencies to which the filter responds, and a rotational symmetric angular Gaussian function that controls the orientation selectivity of the filter [7]:

$$3DG_{ij} = \exp\left(-\frac{\log^2(\omega/\omega_{0i})}{2\times\log^2(\kappa/\omega_{0i})}\right)\times\exp\left(-\frac{\alpha(\phi_j,\theta_j)^2}{2\times\sigma_\alpha^2}\right) \tag{3}$$

The subscripts i and j denote the number of scale and orientations of the filter, respectively. $\alpha(\phi_j, \theta_j)$ is the angle between the azimuth (ϕ_j) and elevation (θ_j) angles. The angular bandwidth is determined by σ_α. κ is a scaling factor used to set the bandwidth of the filter in the radial direction. The scaling of the radial Log-Gabor function is controlled using different wavelengths that are based on multiples of a minimum wavelength, λ_{min}, a user-defined parameter [7]. The filter scale i, and center frequency ω_0 are related as $\omega_{0i}=2/\lambda_{min}\times(\delta)^{i-1}$ where δ is a scaling factor defined for computing the center frequencies of successive filters. In [9], Hacihaliloglu et al. proposed a framework for automating the selection of 2D Log-Gabor filter parameters. However, 2D methods ignore the correlations between adjacent slices and are therefore subject to large spatial compounding errors as well as errors associated with the US beam thickness effects [7]. Therefore, we have extended the automated filter parameter selection approach from slice-based 2D processing to fully 3D.Using this statistical approach we automatically selected all these filter parameters: frequency bandwidth (κ/ω_0), orientation (azimuth (ϕ_j) and elevation (θ_j) angles), scale (λ_{min}) and angular bandwidth (σ_α).The filter bandwidth in the radial direction, $\beta= -2\times(2/ln2)^{0.5}\times ln(\kappa/\omega_0)$, is related to both the speckle   and boundary responses in the US image. To determine the filter frequency bandwidth we analyzed US images with fully developed speckle in elevation direction. The speckle size is estimated as the full-width at half-maximum (*FWHM*) of the autocorrelations. The ratio, κ/ω_0, is computed as $\kappa/\omega_0=\exp(-0.25\times(2\times ln(2))^{0.5}\times FWHM\times r)$. Here r is the pixel size in mm. The strongest response occurs when the 3D Log-Gabor filter is orientated orthogonally to the surface of the bone. The initial filter angles are obtained by clustering the 3D radon transform (*3DRT)* of the 3D B-mode US volume. The projection angles corresponding to the peak values of the *3DRT* represent the angles perpendicular to the high intensity bone response and are chosen as the initial filter angles. These initial angles are used to determine the filter scale (λ_{min}). Since our main interest here is localizing bone contours, which generally appear as ridges in US images, we employ a metric that captures   the 'ridgness' which we calculate as: $3DRS_\gamma=t^{2\gamma}\times(K_1-K_2)^2$. Here t is the scale of the filter ($t= \lambda_{min}$), $K_1=Trace(H(x))+(Trace(H(x))^2-4\times det(H(x)))^{0.5}$, and $K_2=Trace(H(x))-Trace(H(x))^2-4\times det(H(x)))^{0.5}$. *H* is the 3D Hessian matrix where 3D Log-Gabor filter was used during the construction of matrix. We analyze the intensity distribution of *3DRS*$_\gamma$ over all possible scales and select the scale where the sum of intensities is maximized as the correct filter scale. The final filter orientations are obtained by recalculating the *3DRT* for the ridge strength image *3DRS*$_\gamma$, obtained using the optimized filter scale and initial filter orientation, and selecting the maximum value of the *3DRT*. Finally, angular bandwidth (σ_α) is determined to be the peak kurtosis value of the Radon transform of the ridge strength image over multiple angular bandwidth values because a greater kurtosis indicates the variance is due to infrequent high-intensity voxel values. The final bone surface that was used during the registration algorithm is determined as the maximum of the 3D PS value; calculated using optimized filter parameters, along the direction of the US probe.

2.3   Point Cloud Simplification Using Particle Simulation

In this work we employ a modification of the particle simulations (*PSim*) method developed by Pauly et al. [10] to decrease the number of points. The first step in *PSim* is to randomly select n points from the original point cloud, P_S, resulting in a smaller subset of points denoted as P_s. Each point, p_i, in this subset is moved along the force $F_t = F_{in} + F_{ex} + F_j$ and projected back onto the bone surface extracted in Section 2.2finding the nearest neighbor in P_S [10]. F_{in} is the internal force of each point in P_s acting on p_i (4).

$$F_{in} = \sum_{i=1}^{n} k(d - \|p - p_i\|) \times (\frac{p - p_i}{\|p - p_i\|}) \tag{4}$$

Here d defines the size of the neighborhood used to calculate force calculations. *PSim* was originally proposed for closed surfaces [10]; however the bone surfaces have boundaries at the edges of the data. In order to avoid F_{in} forcing points towards low density areas, i.e. the edge of the surface, we have added F_{ex}, implemented as a trough function, which acts as a force to push points away from the edge. Finally, to ensure that points do not overlap we included a small perturbation F_j. After the force has been calculated, each point is moved according to F_t. This step is called the relaxation of the point cloud as described by Pauly et al [10]. These steps are iteratively repeated until equilibrium is reached. The force calculation above was implemented by Pauly et al [10] with an $O(n^2)$ complexity. As this calculation must be computed multiple times in the algorithm, it significantly adds to the overall runtime. We instead optimize the force calculations using the Barnes-Hut algorithm [11], modified to estimate F_t. The Barnes-Hut algorithm first models the point cloud as a hierarchical tree of; where each child node contains no more than one point. The tree is divided into octnode volumes, which are subdivisions of the parent node into eight equal parts. This tree is then used to estimate the distance of clusters of points from the point of interest; if points are sufficiently far away (determined by a constant, θ, and the size of the octnode), the cluster can be modeled as a single point. This algorithm has a complexity of $O(n\log n)$ and reduces the runtime by approximately 50%.

2.4   Experimental Setup and Data Acquisition

Our main clinical interest is to assess pelvis fractures, which include approximately 3% of all skeletal injuries [12]. Over 40% of patients requiring pelvic stabilization suffer long-term complications, usually neurological, urological and non-specific pain [12]. Therefore, the acute management of pelvic ring injuries is complex and demands precise surgery.

Phantom Study: We designed a phantom consisting of a high-resolution radio-opaque Sawbone hemi-pelvis. The iliac crest region of the sawbone pelvis was cut and placed inside an open topped acrylic plastic cylindrical tube. The phantom was suspended in a PVC gel with sixty-one 1mm metal fiducials suspended approximately 1-3 cm from the surface of the gel to ensure any US volume will contain between four and six fiducials (Fig.1.). This setup was scanned using a commercially available ultrasound machine (G.E. Voluson 730, GE Healthcare, Waukesha, WI) using a 3D RSP5-12 transducer. 49 US scans were acquired along the iliac spine of the phantom. The US volumes were 152x198x148 voxels with an isotropic resolution of 0.25mm.

The gold standard surface was provided by a high resolution CT machine (HRpQCT Xtreme CT, Scanco Medical, Wayne, Pennsylvania). The acquired CT volume size was a 482mm×482mm×402mm volume with a resolution of 0.25mm. The fiducials were omitted from registration and were only used for validation purposes.

Clinical Study: The objective of the clinical study was to determine in a live operating room setting whether or not the proposed method could achieve successful registration results with the required accuracy. After obtaining all required ethics approvals, we obtained both CT and US scans from 21 consenting patients admitted to a Level 1 Trauma Centre with pelvic fractures that required a CT scan. The US scans were collected from the iliac crest regions. The voxel resolution for the CT volumes varied between 0.76 mm-0.83 mm in the x and y axes and 1 mm-2 mm in the z axis. We also provide registration accuracy comparison results where local phase features were obtained using empirical *vs* optimized 3D Log-Gabor filter parameters.

(a)   (b)
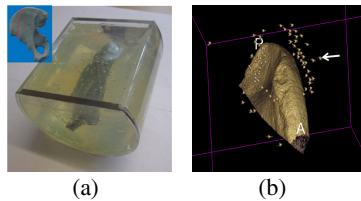
Fig. 1. Phantom validation experiment. (a) Constructed phantom comprised an iliac crest sawbone inside an open-topped acrylic plastic cylindrical tube filled with PVC (cutting plane illustrated in the top left corner). (b) High resolution CT scan of the constructed phantom (white arrow pointing to one of the fiducials used).

3   Results and Discussion

The GMM registration algorithm and *PSim* was implemented in C++, whereas the curvature calculation and optimized 3DPS based bone surface extraction were implemented in MATLAB. All experiments were run on a 2.67GHz Intel(R) Core™i7 central processing unit (CPU) using 64-bit Windows 7 and a 4GB of RAM. The quality of the registration algorithm was evaluated by computing the root mean square error distance between the two registered surface representations, which we will denote as the surface registration error (SRE). We calculated target registration error (TRE) as the average distance between all corresponding fiducial pairs found in both volumes to validate the anatomical accuracy of our proposed method. The number of randomly selected points (n) for *PSim* calculation was chosen as n=500 to provide the optimal trade-off between speed and accuracy. The results of the phantom setup showed an average 0.77 mm (SD 0.29 mm) TRE with maximum TRE of 1.60 mm using the *PSim* point simplification method. The mean SRE for the phantom study using the proposed method was calculated as 0.31mm (SD 0.09 mm). For the naïve down sampling approach [6] the mean TRE increased to 6.84 mm (SD 5.35 mm) with a maximum TRE value of 25.05 mm. The average SRE for naïve down sampling method was 0.28 mm (SD 0.07) (Table 1.). The average run times for *PSim* and naïve down sampling method were 1.40 s (SD 0.30 s) and 1.01s (SD 0.48 s) respectively. The

registration time decreased to 0.97 s (SD 0.55 s) when *n* was set to 300. However, the SRE increased to 0.55 mm (SD 0.24 mm), indicating using 500 points is more robust.

There were no fiducials available for the clinical validation; therefore, only the SRE was measured. The proposed method was able to register all 21 (100%) clinical scans with a mean SRE of 0.72 mm (SD 0.66 mm), while the naïve downsampling approach [6] was able to register only 10 scans (48%). The mean SRE for the naïve downsampling approach was over 3X higher at 2.48 mm (SD 1.66 mm).

Table 1. Quantitative results comparing registration using PSim vs naive downsampling for phantom and clinical scans

	Phantom				Clinical		
	PSim		Naïve [6]		PSim	Naïve [6]	Empirical PS
	SRE (mm)	TRE (mm)	SRE (mm)	TRE (mm)	SRE (mm)	SRE (mm)	SRE (mm)
Mean	0.31	0.77	0.28	6.84	0.72	2.48	1.88
SD	0.09	0.29	0.07	5.35	0.66	1.66	1.75
Median	0.29	0.79	0.27	5.76	0.66	0.92	4
Min	0.15	0.25	0.18	0.80	0.01	0.08	0.32
Max	0.65	1.60	0.48	25.05	2.25	4	4

The success of registration is also affected by the extracted local phase features. Only 12 clinical scans (57%) were successfully registered when using empirically determined filter parameters [7]. The mean SRE for this case was 1.88 mm (SD 1.75 mm). It should be noted that the SRE was lower for all cases, as it is calculated from the extracted outlines which are determined from the 3DPS. Qualitative results of the clinical study can be seen in Figure 2. The US volumes were acquired in a region of the iliac spine unaffected by the fracture. The optimized filter parameters result in the extraction of sharper and more continuous surfaces (Fig. 2c) when compared to empirical method [7] (Fig. 2b). Investigating the registration result we can see that the proposed method successfully aligns the two volumes (Fig. 2d) where a close match between the surfaces is visible compared to the previous method (Fig. 2e).

| (a) | (b) | (c) | (d) | (e) |
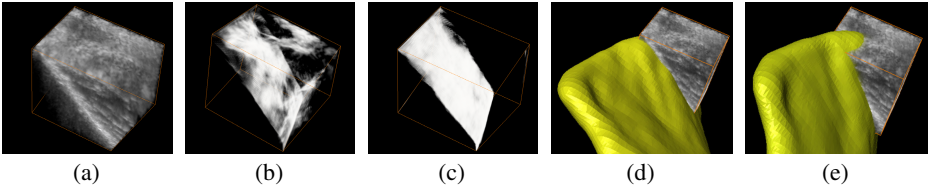
Fig. 2. Qualitative results on clinical data. (a) B-mode US volume obtained by scanning a patient's iliac crest. (b) 3D PS bone surface using empirical filter parameters. (c) Bone surface using our optimized filter parameters. (d) Registration using our proposed method. (e) Registration using naïve down sampling with empirical filter parameters. In (d) and (e) yellow surface is the ROI surface from the iliac crest region.

4   Conclusions

Accurate, robust and fast image registration is essential in US-based image guided interventions in order to provide real-time guidance. The previously proposed methods were either not robust enough to typical US artifacts or could not achieve the

required speed or accuracy to be used in a live operating room setting. We presented a novel method for near real-time 3D US-CT registration in bone imaging deploying a statistical point-based registration in combination with a point cloud optimization method (*PSim*) that better represents the surface. We reduced the run-time by a further 50% by optimizing the force calculations used during point simplification. Furthermore, we incorporated a novel data driven selection of 3D log-Gabor filter parameters in the context of local phase based bone feature extraction in 3D US. The improvements achieved in terms of registration accuracy and robustness compared to the previously proposed state-of-the-art method were up to 89% and 52% for the phantom and clinical data, respectively [6]. Our future work will focus on optimizing the algorithms proposed in this paper to run on a GPU to achieve essentially real-time registration results and on validating our findings with a cadaver study.

References

1. Penney, G.P., Edwards, P.J., King, A.P., Blackall, J.M., Batchelor, P.G., Hawkes, D.J.: A Stochastic Iterative Closest Point Algorithm (stochastICP). In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 762–769. Springer, Heidelberg (2001)
2. Moghari, M.H., Abolmaesumi, P.: Point-Based Rigid-Body Registration Using an Unscented Kalman Filter. IEEE Transactions on Medical Imaging 26(12), 1708–1728 (2007)
3. Penney, G., Barratt, D., Chan, C., Slomczykowski, M., Carter, T., Edwards, P., Hawkes, D.: Cadaver Validation of Intensity-Based Ultrasound to CT Registration. Medical Image Analysis 10(3), 385–395 (2006)
4. Gill, S., Abolmaesumi, P., Fichtinger, G., Boisvert, J., Pichora, D., Borshneck, D., Mousavi, P.: Biomechanically Constrained Groupwise Ultrasound to CT Registration of the Lumbar Spine. Medical Image Analysis (2010) (in press)
5. Khallaghi, S., Mousavi, P., Borschneck, D., Fichtinger, G., Abolmaesumi, P.: Biomechanically Constrained Groupwise Statistical Shape Model to Ultrasound Registration of the Lumbar Spine. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 47–54. Springer, Heidelberg (2011)
6. Brounstein, A., Hacihaliloglu, I., Guy, P., Hodgson, A., Abugharbieh, R.: Towards Real-Time 3D US to CT Bone Image Registration Using Phase and Curvature Feature Based GMM Matching. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 235–242. Springer, Heidelberg (2011)
7. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A.J., Rohling, R.: Bone Segmentation and Fracture Detection in Ultrasound Using 3D Local Phase Features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 287–295. Springer, Heidelberg (2008)
8. Jian, B., Vemuri, B.: Robust Point Set Registration Using Gaussian Mixture Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(8), 1633–1645 (2011)
9. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A., Rohling, R.: Automatic Adaptive Parameterization in Local Phase Feature-Based Bone Segmentation in Ultrasound. Ultrasound in Med. and Biol. 37(10), 1689–1703 (2011)
10. Pauly, M., Gross, M., Kobbelt, L.: Efficient simplification of point- sampled surfaces. In: IEEE Proc. of Visualization (VIS 2002), pp. 163–170 (2002)
11. Barnes, J., Hut, P.: A hierarchical O(N log N) force-calculation algorithm. Nature 324, 446–449 (1986)
12. Petrisora, B., Bhandarib, M.: Injuries to the pelvic ring: Incidence, classification, associated injuries and mortality rates. Current Orthopaedics 19(10), 327–333 (2005)

Hierarchical Attribute-Guided Symmetric Diffeomorphic Registration for MR Brain Images

Guorong Wu, Minjeong Kim, Qian Wang, and Dinggang Shen

Department of Radioloy and BRIC, Univerity of North Carolina at Chapel Hill
{grwu,mjkim,dgshen}@med.unc.edu, qianwang@email.unc.edu

Abstract. Deformable registration has been widely used in neuroscience studies for spatial normalization of brain images onto a standard space. Due to high anatomical variances across individual brains, registration performance could be limited when trying to estimate entire deformation pathway either from template to subject or subject to template. Symmetric image registration offers an effective way to simultaneously deform template and subject images towards each other until they meet at the middle point. Although some intensity-based registration algorithms have nicely incorporated this concept of symmetric deformation, the intensity matching between two images may not necessarily imply the correct matching of anatomical correspondences. In this paper, we integrate both strategies of hierarchical attribute matching and symmetric diffeomorphic deformation for building a new symmetric-diffeomorphic registration algorithm for MR brain images. The performance of our proposed method has been extensively evaluated and further compared with top-ranked image registration methods (SyN and diffeomorphic Demons) on brain MR images. In all experiments, our registration method achieves the best registration performance, compared to all other state-of-the-art registration methods.

1   Introduction

Modern medical imaging technology provides a convenient way to perform clinical diagnosis and research, such as human brain development, aging, and disease-related abnormalities. In order to measure subtle anatomical or functional difference, accurate deformable registration plays an important role in dealing with confounding intra-subject variability in the longitudinal studies and inter-subject variability in the cross-sectional studies.

A lot of deformable image registration methods have been proposed in the last decades. A comprehensive survey can be found in [1]. In general, the goal of deformable registration is to estimate the deformation field from template (fixed image) to subject (moving image) by maximizing a certain image similarity measurement between template and warped subject image. As a highly ill-posed problem, regularizing the deformation field is necessary for resolving the uncertainty in correspondence detection during the registration.

Recently, 14 deformable registration methods have been comprehensively evaluated in [2] based on the registration performance on human brain MR images.

Interestingly, all of these 14 registration methods are intensity-based. However, the use of image intensity only is not sufficient to deal with the confounding inter-subject variations, due to the lack of specific features to establish reasonable anatomical correspondences. In [2], SyN [3] is recommended as the top deformable registration method, which uses the concept of symmetric registration by estimating the deformation pathway from two ends (template and subject) to the middle point.

Inspired by this symmetric registration mechanism [3], we propose a feature-based symmetric registration method in this paper with three novel strategies. First, we define attribute vectors for detecting anatomical correspondences and key points with distinctive image feature for hierarchically driving the entire registration procedure in image registration. Specifically, the attribute vector on each voxel considers not only local image appearance but also edge information. The key points are randomly sampled in the non-uniform and hierarchical fashion, assuring that they have distinctive features and cover the entire brain. Second, robust correspondence matching (by comparing attribute vectors) is performed in a symmetric way. We simultaneously optimize the deformations on the key points from both template and subject until they meet at the middle point of the whole deformation pathway (between template and subject). Since the correspondence matching is always performed between the key points on the deformed template and the deformed subject in the progress of image registration, the correspondence matching is relatively easier, compared to the case of directly finding correspondences between the original template and subject images. Hence, our method is able to handle large anatomical variations among individual brains under registration. Third, we adapt the registration procedure to a space of diffeomorphic transformation by following the efficient log-domain approach in [4]. Therefore, the estimated deformation field by our registration method is smooth and invertible, but the computation time can be greatly reduced, compared with other complicated diffeomorphic registration methods [3].

To compare the registration performance of our proposed registration method, we choose two state-of-the-art registration methods, i.e., SyN [3] and diffeomorphic Demons [4], as the references, since they achieve the top-ranked performance as evaluated in [2] and are widely used in neuroscience study. The registration accuracy is comprehensively evaluated on real human brain MR images (elderly brains aged from 65-85, LONI LBPA40 [5], and NIREP NA0 [6] datasets), all with manually-labeled ROIs. In all experiments, our proposed feature-based registration method achieves the best registration performance in terms of registration accuracy.

2   Methods

The goal of deformable registration is to find a transformation $F = \{f(x) | f(x) = x + h(x), \ x \in \Re^3\}$ between template T and subject S, with displacement $h(x)$ defining the mapping of the point coordinates Ω_T of template T to Ω_S of subject S. Here, $\Omega_T, \Omega_S \subset \Re^3$. In this section, we will first present the overview of our registration method in Section 2.1. The energy function and its solution will be explained in Section 2.2 and Section 2.3, respectively.

2.1   Overview of Our Symmetric Feature-Based Registration Method

The overview of our symmetric feature-based registration method is shown in Fig. 1. Instead of estimating the deformation field F from template to subject, we estimate two deformation fields starting from template and subject, respectively. Thus, we can obtain the deformation fields ϕ_1 from template space and ϕ_2 from the subject. Both ϕ_1 and ϕ_2 point towards the common space in the middle of the pathway between template and subject. In the end, the deformation field from template to subject can be calculated by $F = \phi_1 \circ \phi_2^{-1}$, where $'\circ'$ denotes the composition of two deformation fields. Similarly, the inverse deformation field F^{-1} from subject to template can be obtained by $F^{-1} = \phi_2 \circ \phi_1^{-1}$.

As mentioned earlier, anatomical correspondence is very important in deformable image registration. Therefore, we define attribute vector $\vec{a}$ on each voxel to establish the correspondence. To further improve the correspondence detection, we hierarchically select the key points (located at the salient regions in the MR brain images) to steer the deformation of registration, as well as to guide the deformation of other less-distinctive points. As displayed in Fig. 1(a), only a small number of points located at the critical areas of brain image are selected as the key points (red points denote for key points of template and blue points for key points of subject) to drive the image registration. With the progress of registration as in the bottom of Fig. 1(a), more and more voxels are qualified as the key points to refine the deformation field, as it can seen in the bottom of Fig. 1(a) that ϕ_1 and ϕ_2 change from global deformation in the beginning to local deformation in the end of registration. The definition of attribute vector and hierarchical key point selection will be explained next.

(a) Overview of our symmetric registration method   (b) Non-uniform sampling
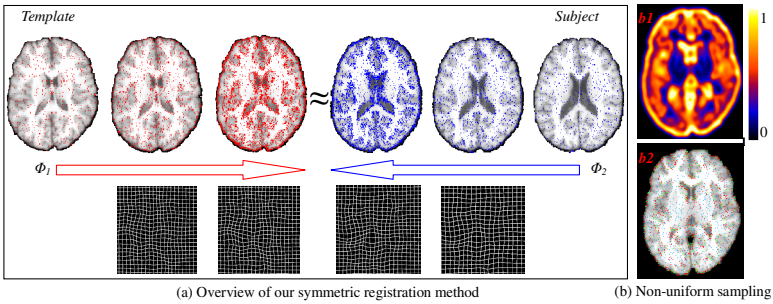
Fig. 1. The framework of our symmetric feature-based registration algorithm (a) and the key modules in our method, i.e., key point selection by non-uniform sampling (b).

2.2   Energy Function in Our Symmetric Registration Method

In our symmetric registration method, both template and subject will deform toward the middle point iteratively, as shown in Fig. 1. (a). Thus, we use $T^k = T(\phi_1^k)$ and $S^k = S(\phi_2^k)$ to denote the tentatively-warped template and subject images w.r.t. the deformation fields ϕ_1^k and ϕ_2^k in the end of the kth iteration, respectively.

<u>Attribute Vector:</u> Attribute vector is used as the morphological signature to characterize geometric information around each voxel. Although some rich image descriptors, e.g.,

SIFT and SURF, work well on correspondence matching, they are computational expensive and sensitive to the structural variations across individual brain images. For robust correspondence matching, we incorporate both local image appearance and edge information in our registration method by arranging all image intensities and gradients in a neighborhood (with radius r, e.g., $r = 3$mm) into the attribute vector $\vec{a}$. Normalized cross correlation is used as the similarity measure to evaluate the similarity between point x in the deformed template T^k and point y in the deformed subject S^k, which is denoted as $NCC(\vec{a}_T^k(x), \vec{a}_S^k(y))$. Furthermore, we define the feature discrepancy as $\eta(\vec{a}_T^k(x), \vec{a}_S^k(y)) = \frac{1 - NCC(\vec{a}_T^k(x), \vec{a}_S^k(y))}{2}$, which ranges from 0 to 1. Apparently, we evaluate the similarity not only on the single pair of points x and y, but also on their respective sub-volumes. Since we explicitly detect correspondences only on the key points, not every voxel in the image, our registration method is efficient, although the computation of normalized cross correlation is more expensive than simple intensity difference.

Key Points Selection: Here, we use an importance sampling strategy to hierarchically select the key points in the image. Specifically, we first smooth, and normalize the gradient magnitude values in the whole image domain. Then, these values are roughly considered as the importance of each voxel during the registration. It is worth noting that the gradient magnitude is selected as the importance of each voxel mainly because of its low computational cost, although other advanced methods can be also applied here. Based on the importance map, a set of key points can be sampled, with higher importance value indicating the higher likelihood of the underlying location being selected in the non-uniform sampling. Fig. 1 shows the non-uniform sampling based on the importance map (Fig. 1(b1)). The initial subset of key points and also the key points selected in the later two stages of registration are displayed in blue, green, and red in Fig. 1(b2), respectively. It is clear from Fig. 1(b2) that the key points are more concentrated at areas with rich edge information, where the importance values are high. After we apply the non-uniform sampling to both T^k and S^k, we can obtain the template key points $X^k = \{x_i^k | i = 1, \dots, M_T^k\}$ and the subject key points $Y^k = \{y_i^k | i = 1, \dots, M_S^k\}$ at the kth iteration of registration, where M_T^k and M_S^k are the numbers of key points in the deformed template and deformed subject in the kth iteration.

Energy Function: In $(k + 1)^{th}$ iteration, we will estimate the incremental deformation φ_1^{k+1} from T^k to S^k and φ_2^{k+1} from S^k to T^k. As we will explain later, the refined deformation fields ϕ_1^{k+1} and ϕ_2^{k+1} in the end of the $(k + 1)^{th}$ iteration can be obtained by integrating the incremental deformation fields φ_1^{k+1} and φ_2^{k+1}.

Given the currently estimated incremental deformation $\varphi_1^{k+1}(x_i^k)$ at key point x_i^k, exhaustive search is performed to refine the correspondence w.r.t. each candidate point u in a search neighborhood $n\left(\varphi_1^{k+1}(x_i^k)\right)$, according to the two criteria: 1) the feature discrepancy should be as small as possible between $\vec{a}_T^k(x_i^k) \in \Omega_{T^k}$ and $\vec{a}_S^k(u) \in \Omega_{S^k}$; 2) the spatial distance between candidate point u and the tentatively estimated location $\varphi_1^{k+1}(x_i^k)$, i.e., $\left\| \varphi_1^{k+1}(x_i^k) - u \right\|^2$, should be as close as possible.

Since there are a lot of uncertainties in correspondence matching, encouraging multiple correspondences is proven effective to alleviate the ambiguity issue. For a particular template key point x_i^k, a probability π_i^u (called as spatial assignment) is

assigned to each candidate point $u \in n(\varphi_1^{k+1}(x_i^k))$ during correspondence matching. For robust correspondence matching, the candidate points even with large discrepancy still might have the chance to contribute to the correspondence matching in the beginning of registration. As the registration progresses, only the candidates with the most similar attribute vectors will be considered until the exact one-to-one correspondence is allowed in the end of registration for achieving the registration specificity. This dynamic procedure can be encoded with the entropy term on the probability, i.e., $\pi_i^u \cdot log\pi_i^u$. Here, high degree of entropy implies the fuzzy assignment while low degree means almost binary matching. We use a scalar value t^{k+1} to act as the temperature to enforce the dynamic change on correspondence assignment.

The optimization of incremental deformation field φ_2^{k+1} follows the same way as discussed above, except we use τ_j^v to denote the assignment of each candidate point v in the search neighborhood $n\left(\varphi_2^{k+1}(y_j^k)\right)$. Thus, the total energy function in estimating incremental deformation field φ_1^{k+1} and φ_2^{k+1} is given as:

$$
\begin{aligned}
E(\varphi_1^{k+1}, \varphi_2^{k+1}) = &\sum_{i=1}^{M_T^k} \sum_{u \in n\left(\varphi_1^{k+1}(x_i^k)\right)} \left\{ \pi_i^u \cdot \left[\eta\left(\vec{a}_T^k(x_i^k), \vec{a}_S^k(u)\right) + \|\varphi_1^{k+1}(x_i^k) - u\|^2 \right] + t^{k+1} \cdot (\pi_i^u \cdot log\pi_i^u) \right\} \\
&+ \sum_{j=1}^{M_S^k} \sum_{v \in n\left(\varphi_2^{k+1}(y_j^k)\right)} \left\{ \tau_j^v \cdot \left[\eta\left(\vec{a}_T^k(v), \vec{a}_S^k(y_j^p)\right) + \|\varphi_2^{k+1}(y_j^k) - v\|^2 \right] + t^{k+1} \cdot (\tau_j^v \cdot log\tau_j^v) \right\} \\
&+ \beta \cdot \{B(\varphi_1^{k+1}) + B(\varphi_2^{k+1})\},
\end{aligned} \tag{1}
$$

where $B(.)$ measures the bending energy of incremental deformation fields [7]. β in Eq. 1 is a scalar value to control the strength on deformation smoothness.

2.3   Optimization of Symmetric Deformation Pathways

First, the spatial assignment π_i^v can be calculated by minimizing E in Eq. 1 w.r.t. π_i^u:

$$
\pi_i^u = \exp\left\{ -\frac{\eta\left(\vec{a}_T^k(x_i^k), \vec{a}_S^k(u)\right) + \left\|\varphi_1^{k+1}(x_i^k) - u\right\|^2}{t^{k+1}} \right\}. \tag{2}
$$

Similarly, the spatial assignment τ_j^v can be obtained as:

$$
\tau_j^v = \exp\left\{ -\frac{\eta\left(\vec{a}_T^k(v), \vec{a}_S^k\left(y_j^p\right)\right) + \left\|\varphi_2^{k+1}\left(y_j^k\right) - v\right\|^2}{t^{k+1}} \right\}. \tag{3}
$$

It is clear that spatial assignment π_i^u and τ_j^v are penalized in the exponential way according to the discrepancy degree η. Notice that the temperature t^{k+1} is the denominator of the exponential function in Eqs. 2 and 3. Therefore, when t^{k+1} is very high in the beginning of registration, even though the discrepancy η might be large or the candidate location is far away, the candidate point still might have the chance to be selected as the correspondence. As registration progresses, the specificity of correspondence will be encouraged by gradually decreasing the temperature t^{k+1} to a small degree, until only the candidate point with the smallest discrepancy being selected as the correspondence in the end of registration.

After obtaining π_i^u for each candidate u, the estimated incremental deformation on key point x_i^k can be computed by optimizing energy function E in Eq. 1 w.r.t. φ_1^{k+1}:

$$\hat{\varphi}_1^{k+1}(x_i^k) = \sum_{u \in n(\varphi_1^{k+1}(x_i^k))} \pi_i^u \cdot u \Big/ \sum_{u \in n(\varphi_1^{k+1}(x_i^k))} \pi_i^u. \tag{4}$$

Similarly, the incremental deformation of each y_j^k can be updated as:

$$\hat{\varphi}_2^{k+1}(y_j^k) = \sum_{v \in n(\varphi_2^{k+1}(y_j^k))} \tau_j^v \cdot v \Big/ \sum_{v \in n(\varphi_2^{k+1}(y_j^k))} \tau_j^v. \tag{5}$$

The estimated incremental deformation fields $\hat{\varphi}_1^{k+1}$ and $\hat{\varphi}_2^{k+1}$ are sparse which have the displacements only on the key points. Then TPS [7] is used to immediately inter-polate the dense deformation fields $\hat{\varphi}_1^{k+1}$ and $\hat{\varphi}_2^{k+1}$ with the bending energy mini-mized. In order to ensure the invertibility of output deformation fields, we follow an efficient non-parametric diffeomorphic approach [4] to adapt the optimization of ϕ_1^{k+1} and ϕ_2^{k+1} to the space of diffeomorphic transformation. The basic idea is to consider the incremental deformation fields $\hat{\varphi}_1^{k+1}$ and $\hat{\varphi}_2^{k+1}$ in the vector space of velocity fields and then map them to the space of diffeomorphism through the expo-nentials, i.e., $\exp(\hat{\varphi}_1^{k+1})$ or $\exp(\hat{\varphi}_2^{k+1})$.

Specifically, the following steps will be applied to calculate the deformation field $\hat{\phi}_1^{k+1}$ under the framework of diffeomorphism: 1) compute the exponential of incre-mental deformation field $\exp(\hat{\varphi}_1^{k+1})$ by the scaling and squaring method [4]; 2) compose the exponential with the previously estimated deformation field by $\hat{\phi}_1^{k+1} = \phi_1^k \circ \exp(\hat{\varphi}_1^{k+1})$; 3) the inverse deformation field can be computed by $(\hat{\phi}_1^{k+1})^{-1} = (\phi_1^k)^{-1} \circ \exp(-\hat{\varphi}_1^{k+1})$. Similarly, the deformation pathway $\hat{\phi}_2^{k+1}$ and its inverse $(\hat{\phi}_2^{k+1})^{-1}$ can be computed by performing these three steps.

Finally, the latest estimated deformation pathway $\hat{\phi}_1^{k+1}$ and $\hat{\phi}_2^{k+1}$ will be used to deform the original template and subject images, and obtain T^{s+1} and S^{k+1} for the next iteration. The whole procedure (i.e., estimating the incremental deformation fields (φ_1^{k+1} and φ_2^{k+1}) and updating the symmetric deformation pathways ϕ_1^{k+1} and ϕ_2^{k+1}) will be repeated until the template and subject images meet at the middle point.

3   Experiments

To evaluate the registration performance, we test our symmetric feature-based regis-tration method on elderly brain images, as well as NIREP-NA0 and LONI-LPBA40 datasets, by comparing it with the two state-of-the-art registration methods, i.e., SyN [3] and diffeomorphic Demons (D. Demons) [4], which demonstrate good perfor-mance in [2]. For SyN and diffeomorphic Demons registration methods, we use their best parameters as listed in the supplementary document of [2].

Registration Results on 18 Elderly Brains: In this experiment, 18 MR brain images of elderly subjects were used, each with image dimension $256 \times 256 \times 124$ and resolution $0.9375 \times 0.9375 \times 1.5mm^3$. After selecting one subject as the template, we registered all other 17 subjects onto the template by SyN, diffeomorphic Demons, and our proposed method, respectively. The mean images (averaging up all registered

(a) Mean image by SyN

(b) Variance image by SyN

(c) Mean image by diffeomorphic demons

(d) Variance image by diffeomorphic demons

(e) Mean image by S-HAMMER

(f) Variance image by S-HAMMER
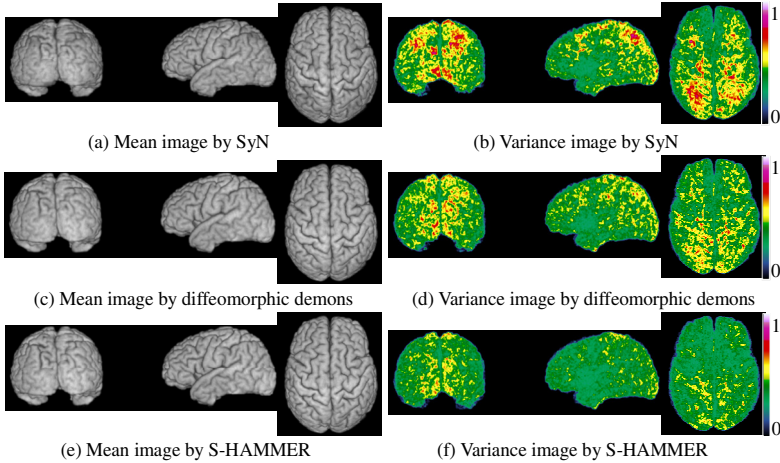
Fig. 2. Demonstration of registration performance on 18 elderly MR brain images. The mean images upon all registered images by SyN, diffeomorphic demons, and our method are shown in (a), (c), and (e), respectively. To examine the registration consistency across different subjects, the local entropy of tissue probability maps across all aligned images by SyN, diffeomorphic Demons, and our method are shown in (b), (d), and (f), respectively, indicating that our method has the least variation.

subject images) by SyN, diffeomorphic Demos, and our registration method are shown in Fig. 2(a), (c), and (e), respectively.

Since the images have been segmented into GM, WM, CSF and VN, we can calculate the voxel-wise overlap ratio between the warped subjects and the template (Table 1). Here, we set the target overlap ratio as the overlap measurement since it was used in [2] to evaluate all 14 registration methods. By masking out the background voxels, the averaged target overlap ratios between template and each aligned subject are 70.10% for SyN, 70.88% by deformorphic Demons, and 74.13% by our proposed method, respectively. For further quantification of registration consistency, we computed the voxel-by-voxel entropy of tissue probability maps across all aligned subjects. High entropy value indicates less consistency. For our proposed method, the averaged entropy is 0.40, which is nearly 17% improvement compared to 0.48 obtained by deformorphic Demons, and 35% improvement compared to 0.62 obtained by SyN. The respective 3D map of entropy value is shown in Fig. 2 by SyN (2b), deformorphic Demons (2d), and our method (2f). It is worth noting that the improvement on tissue overlap ratios between our method and other two methods are significant in paired t-test ($p<0.05$).

Table 1. Target overlap ratio of white matter, gray matter, ventricle on 18 elderly brains by SyN, diffeomorphic Demons (D. Demons), and our method. (unit: %)

	White matter	Gray matter	Ventricle	Overall
SyN	71.28±3.33	56.63±3.58	82.40±3.35	70.10±3.42
D. Demons	72.05±5.24	58.22±8.00	82.36±4.50	70.88±6.34
Our method	75.22±3.22	62.36±3.92	84.82±1.86	74.13±2.86

NIREP NA0 Dataset: One of the 16 images is randomly selected as the template. Then we register the rest 15 images to the template image by SyN, diffeomorphic Demons, and our proposed method. After registration, the target overlap ratio can be calculated for each of the 32 manually delineated labels. We repeat this procedure for 5 times by selecting different subject as template. The mean and standard deviation of target overlap ratios across all subjects and all ROIs are 60.25±1.67% by SyN, 64.23±1.38% by diffeomorphic Demons, and 65.17±1.16% by our registration method, respectively.

LONI LPBA40 Dataset: In this experiment, we use the LONI LPBA40 dataset [5] with 40 brain images and 54 manually labeled ROIs in each brain image. Similarly, we employ SyN, diffeomorphic Demons, and our method to align 39 brain images to a randomly selected template image. We repeat this procedure for 5 times as well. The overall tissue overlap ratio is 73.23±1.35% by SyN, 73.12±1.62% by diffeomorphic Demons, and 74.00±1.26% by our method, respectively.

4   Conclusion

In this paper, we proposed a new feature-based symmetric registration method for MR brain images. Compared with other intensity-based registration methods, our method achieves more accurate registration. Our symmetric feature-based registration method produces better registration results than the two top-ranked intensity-based registration methods (SyN and D. Demons) on all three experiments.

References

1. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis 2, 1–36 (1998)
2. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., Song, J.H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R.P., Mann, J.J., Parsey, R.V.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage 46, 786–802 (2009)
3. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26–41 (2008)
4. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: efficient non-parametric image registration. NeuroImage 45, S61–S72 (2009)
5. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39, 1064–1080 (2008)
6. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the Non-rigid Image Registration Evaluation Project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
7. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, 567–585 (1989)

Uncertainty-Based Feature Learning
for Skin Lesion Matching
Using a High Order MRF Optimization Framework

Hengameh Mirzaalian[1], Tim K. Lee[1,2,3], and Ghassan Hamarneh[1]

[1] Medical Image Analysis Lab, Simon Fraser University
[2] Cancer Control Research, BC Cancer Agency
[3] Department of Dermatology and Skin Science, University of British Columbia
{hma36,hamarneh}@sfu.ca, tlee@bccrc.ca

Abstract. We formulate the pigmented-skin-lesion (PSL) matching problem as a relaxed labeling of an association graph. In this graph labeling problem, each node represents a mapping between a PSL from one image to a PSL in the second image and the optimal labels are those optimizing a high order Markov Random Field energy (MRF). The energy is made up of unary, binary, and ternary energy terms capturing the likelihood of matching between the points, edges, and cliques of two graphs representing the spatial distribution of the two PSL sets. Following an exploration of various MRF energy terms, we propose a novel entropy energy term encouraging solutions with low uncertainty. By interpreting the relaxed labeling as a measure of confidence, we further leverage the high confidence matching to sequentially constrain the learnt objective function defined on the association graph. We evaluate our method on a large set of synthetic data as well as 56 pairs of real dermatological images. Our proposed method compares favorably with the state-of-the-art.

1 Introduction

The presence of a large number of pigmented skin lesions (PSL) is a strong predictor of malignant melanoma [7]. Since detecting newly appearing, disappearing, and changing PSL is important for early detection of the disease, many dermatologists advocate total-body photography for high-risk patients (Figure 1(a)). However, manual inspection and matching of PSLs is a subjective, tedious, and error prone task. A computer vision system for tracking the corresponding PSLs greatly improves the matching process, thereby easing the workload on dermatologists while also improving matching accuracy and removing operator variability [7]. There exists limited works on automating the matching between lesions. Huang and Bergstresser developed a PSL matching algorithm based on a Voronoi decomposition of the image space [3]. Yet, their method does not deal with the presence of the newly appearing or disappearing PSLs. Prednia and White performed affine registration between the two sets of PSLs [8]. However, their method does not take into account the elastic deformation of the human back. Roning and Riech defined a set of geometric properties as a similarity metric to find the corresponding PSLs. Their method requires manually determining two initial matches

(a)   (b)   (c)   (f) With learning
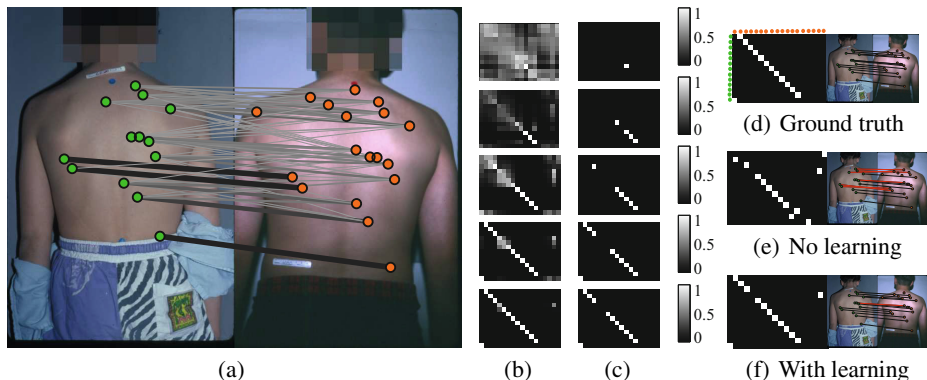
Fig. 1. (a) Example back images of the same subject at two different times. The green and red dots are overlaid at the PSL's coordinates. The thickness and the color of the edges encode the matching confidence between the connected points; the thicker and darker the line, the higher the confidence (please refer to Section 2.3 for more details). The five rows in (b) and (c) represent the output of five iterations of the learning step. (b) The probabilistic solutions. (c) The selected high confidence matchings. The ground truth matching is shown in (d). (e) and (f) show the estimated matching without and with the learning step. Wrong matches are shown in red on the back images. It can be noticed that the unsupervised learning step improves the matching accuracy (i.e. less red lines).

[9]. The authors in [7] computed the matching probabilities of the edges of two graphs representing the spatial distribution of the two PSL sets. They then extracted pointwise probabilities utilizing the marginalization matrix of the computed pairwise matchings. However, they did not make use of high-order term to the PSL matching. Recently, there have been several works on high order graph matching, combining both appearance similarity and geometric compatibility [1,10,11,12].

Compared with the previous works on PSL matching, we present the first application of high-order term to PSL matching. Our approach is most closely related to the work of Zeng et. al [12], who formulate a non-rigid surface registration problem as a high order graph matching problem and extract the matchings by solving a corresponding pseudo-boolean function. Their matching cost function depends on the feature appearance and geometric compatibility of the pair-wise and triplet-wise correspondences (Section 2.1). To solve their non-convex optimization problem, they make use of the dual-decomposition (DD) approach, similar to the work by Torresani et. al [10]. Our method differs from those in [10,12] in several ways. First, we relax the labels to continuous variables. By interpreting the relaxed labeling as a measure of confidence, we sequentially leverage the high confidence matchings via a self-learning approach to learn the features of the association graph (Section 2.3). We further propose to add a novel entropy energy term encouraging solutions with low uncertainty. We evaluate our method on a large set of synthetic data (hundreds of pairs) as well as 56 pairs of real dermatological images. The experimental results confirm the usefulness of adopting the entropy term and the unsupervised learning procedure (Section 3).

(a) $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathbb{G}$ (b)   (c)   (d)
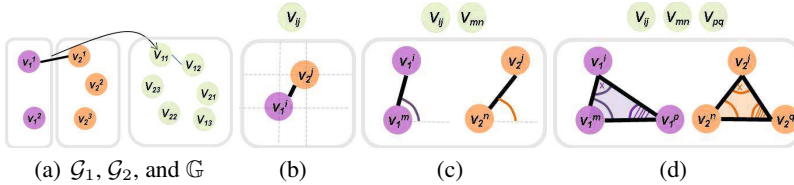
Fig. 2. (a) Association graph $\mathbb{G}$ for graph matching between two graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$. Each node in $\mathbb{G}$ represents a connection between a point in $\mathcal{G}_1$ and a point in $\mathcal{G}_2$. The matching problem between $\mathcal{G}_1$ and $\mathcal{G}_2$ is formulated as a labeling problem for $\mathbb{G}$. (b-d) illustrate examples of the unary, binary, and ternary terms used in the MRF-based labeling cost function in (1) (please refer to Section 2.1 for further details).

2 Method

Let us denote the PSLs coordinates of the lth image by a graph $\mathcal{G}_l(\mathcal{V}_l, \mathcal{E}_l, \mathcal{C}_l)$, $l \in \{1, 2\}$, consisting of a set of nodes $\mathcal{V}_l$ ($|\mathcal{V}_l| = N_l$), edges $\mathcal{E}_l \subset \mathcal{V}_l \times \mathcal{V}_l$, and cliques $\mathcal{C}_l \subset \mathcal{V}_l \times \mathcal{V}_l \times \mathcal{V}_l$. We define a set of intra and inter-edges between the graphs to encode features related to the nodes connected by the edges and the cliques. An intra edge $\mathcal{E}_{lm,ln} \in \mathcal{G}_l$ connects the mth vertex $\mathcal{V}_{lm}$ to the nth vertex $\mathcal{V}_{ln}$, where $n \neq m$. An inter-edge $\mathcal{E}_{1m,2n}$ connects $\mathcal{V}_{1m}$ to $\mathcal{V}_{2n}$. Our aim is to find a mapping $\Pi(\mathcal{V}_1) \rightarrow \mathcal{V}_2$.

The matching problem (i.e. finding the mapping Π) can be formulated as a graph labeling problem. To this end, given $\mathcal{G}_l|_{l=1,2}$, we first construct their association graph $\mathbb{G}(\mathbb{V}, \mathbb{E}, \mathbb{C})$, in which each vertex in $\mathbb{V}$ corresponds to an inter-edge, e.g. $\mathbb{V}_{1m,2n} = \mathbb{V}_{mn} \leftrightarrow \mathcal{E}_{1m,2n}$ ($|\mathbb{V}| = N_1 N_2$) (Figure 2(a)). The matching problem can then be solved by binary labeling, x, of $\mathbb{G}$ [10]. A correspondence $\Pi(\mathcal{V}_{1m}) \rightarrow \mathcal{V}_{2n}$ is active iff $x(\mathbb{V}_{mn}) = 1$ and 0 otherwise. The details describing the objective function for binary labeling is provided in Section 2.1.

Compared with [10], we solve the matching problem as a relaxed (fuzzy) labeling, i.e. $x \in [0, 1]$. We interpret the fuzzy labels as a measure of confidence. The high confidence matchings are then extracted for unsupervised learning of the features of the association graph (Section 2.3).

Let us denote the label by $\ell \in \{0, 1\}$. Then, $x^\ell(\mathbb{V})$ represents our confidence in $\mathbb{V}$ having the label ℓ. Since we have the following equality $x^0(\mathbb{V}) = 1 - x^1(\mathbb{V})$ in our framework, we denote x^1 by x for simplicity.

2.1 MRF-Based Binary Labeling

MRF-optimization seeks the labeling x_p for each vertex $\mathbb{V}_p$ of graph $\mathbb{G}(\mathbb{V}, \mathbb{E}, \mathbb{C})$ by optimizing an energy function of the form:

$$E(x) = w_u \sum_{p \in \mathbb{V}} \phi_x(x_p) + w_b \sum_{(p,q) \in \mathbb{E}} \phi_{xx}(x_p, x_q) + w_t \sum_{(p,q,z) \in \mathbb{C}} \phi_{xxx}(x_p, x_q, x_z) \quad (1)$$

where ϕ_x is the unary term which measures the likelihood of labeling a vertex with a specific label disregarding the labels of any of the neighbours; and ϕ_{xx} and ϕ_{xxx} are

regularization terms penalizing different label configurations of neighboring vertices. w's are the weights of the different terms.

We define our unary term as a weighted sum of the two energy terms:

$$\phi_x(x_{ij}) = w_u^{(1)}\phi_x^{(1)}(x_{ij}) + w_u^{(2)}\phi_x^{(2)}(x_{ij}). \tag{2}$$

$\phi_x^{(1)}(x_{ij})$ measures the dissimilarity between the appearance descriptors of $\mathcal{V}_{1i}$ and $\mathcal{V}_{2j}$, denoted by $\mathcal{F}_{\mathcal{V}_{1i}}$ and $\mathcal{F}_{\mathcal{V}_{2j}}$:

$$\phi_x^{(1)}(x_{ij}) = x_{ij}d_u(\mathcal{V}_{1i}, \mathcal{V}_{2j}) + (1 - x_{ij})(1 - d_u(\mathcal{V}_{1i}, \mathcal{V}_{2j})), \tag{3}$$

$$d_u(A, B) = \sum_{r=1}^{R} |\mathcal{F}_A(\alpha_r) - \mathcal{F}_B(\beta_r)|, \quad \alpha = [\alpha_1\alpha_2...\alpha_R], \beta = [\beta_1\beta_2...\beta_R]$$

α, β are the indices of $\mathcal{F}_A$ and $\mathcal{F}_B$, which are compared to each other in computing d_u in (3), and are given by:

$$\mathcal{F}_{\mathcal{V}_{lm}} = \mathbf{x}(\mathcal{V}_{lm})\mathbf{1}_{1 \times N_l+1} - [\mathbf{0}, \mathbf{x}(\mathcal{V}_{l1}), \mathbf{x}(\mathcal{V}_{l2}), ..., \mathbf{x}(\mathcal{V}_{lN_l})], \quad l \in \{1, 2\}. \tag{4}$$

where $\mathbf{x}_{2 \times 1}$ is the normalized coordinate of the PSLs resulting from applying the skin back-template proposed in [7]. In the first iteration, α and β in (3) are initialized with 1. Therefore, d_u measures the Euclidean distance between $\mathcal{V}_{1i}$ and $\mathcal{V}_{2j}$, i.e. $d_u\Big(\mathcal{F}_{\mathcal{V}_{1i}}(\alpha = 1), \mathcal{F}_{\mathcal{V}_{2j}}(\beta = 1)\Big) = |\mathbf{x}(\mathcal{V}_{1i}) - \mathbf{x}(\mathcal{V}_{2j})|$, and later on, as explained in Section 2.3, α and β in (3) will be updated in a sequential learning step to include more entries of $\mathcal{F}$ in computing d_u.

$\phi_x^{(2)}(x_{ij})$ in (2) is our new *entropy* term, which is used to encourage the cost function towards solutions with low entropy or low uncertainty:

$$\phi_x^{(2)}(x_{ij}) = -\Big(x_{ij}\log_2 x_{ij} + (1 - x_{ij})\log_2(1 - x_{ij})\Big) \approx x_{ij}(1 - x_{ij}). \tag{5}$$

Equation (5) shows a quadratic approximation term achieved using a second order Taylor expansion. We treat x as a probability when calculating Shannon's entropy although we didn't present our method in a formal probabilistic framework. Nevertheless, the intuition of having higher uncertainty as x nears 0.5 and lower uncertainty as x gets close to 1 or 0 still holds.

To measure compatibility between pairwise correspondences, we use:

$$\phi_{xx}(x_{ij}, x_{mn}) = x_{ij}x_{mn}d_b(\overrightarrow{\mathcal{V}_{1i}\mathcal{V}_{1m}}, \overrightarrow{\mathcal{V}_{2j}\mathcal{V}_{2n}}) + (1 - x_{ij}x_{mn})(1 - d_b(\overrightarrow{\mathcal{V}_{1i}\mathcal{V}_{1m}}, \overrightarrow{\mathcal{V}_{2j}\mathcal{V}_{2n}}))$$

$$\overrightarrow{\mathcal{V}_{li}\mathcal{V}_{lm}} = \mathbf{x}(\mathcal{V}_{li}) - \mathbf{x}(\mathcal{V}_{lm}), \quad d_b(\overrightarrow{A}, \overrightarrow{B}) = \omega_b^1|1 - \frac{\overrightarrow{A}.\overrightarrow{B}}{|\overrightarrow{A}||\overrightarrow{B}|}| + \omega_b^2||\overrightarrow{A}| - |\overrightarrow{B}||. \tag{6}$$

d_b evaluates the length and direction agreement between the line segments $\overrightarrow{A}$ and $\overrightarrow{B}$. w_b^1, w_b^2 weight the direction and length terms.

To measure the compatibility in corresponding triplets, e.g. triangles $\mathcal{T}_1 = \mathcal{V}_{1i}\widehat{\mathcal{V}_{1m}}\mathcal{V}_{1p}$ and $\mathcal{T}_2 = \mathcal{V}_{2j}\widehat{\mathcal{V}_{2n}}\mathcal{V}_{2q}$, we use:

Algorithm 1.1. Our proposed MRF-based point matching algorithm.

1: **Input**: Two point sets $\mathcal{V}_1$ and $\mathcal{V}_2$; the spatial coordinates of the points: $\mathbf{x}(\mathcal{V}_1)$ and $\mathbf{x}(\mathcal{V}_2)$.

2: **Output**: A mapping between the vertices: $\Pi(\mathcal{V}_1) \rightarrow \mathcal{V}_2$.

3: **Initialization**: Construct $\mathbb{G}(\mathbb{V}, \mathbb{E})$ (Section 2); construct $\mathcal{F}_{\mathcal{V}_{lm}}$ (4); $\tau = 0.9$; $\alpha = 1$; $\beta = 1$;
      $w_u^{(1)} = 0.04$; $w_u^{(2)} = 0.1$; $w_b^{(1)} = w_b^{(2)} = 0.04$; $w_t^{(1)} = w_t^{(2)} = 0.02$.

4: Compute d_b (6), d_t (7), and $d_u\Big(\mathcal{F}_{\mathcal{V}_{1i}}(\alpha = 1), \mathcal{F}_{\mathcal{V}_{2j}}(\beta = 1)\Big) = |\mathbf{x}(\mathcal{V}_{1i}) - \mathbf{x}(\mathcal{V}_{2j})|$.

5: Define the objective function $E(x) = \mathbf{func}(\phi_x^{(1)}, \phi_x^{(2)}, \phi_{xx}, \phi_{xxx})$ (Section 2.1).

6: Optimize $E(x)$.   ▷ e.g. apply SP [11] or TIP [1] to maximize $x = \max_x E(x)$.

7: $(\mathcal{A}, \mathcal{B}) = \{(i, j) | x_{ij} > \tau\}$.   ▷ $\mathcal{A}$ and $\mathcal{B}$ are the indices of the high confidence nodes in $\mathbb{G}$.

8: **if** $\mathcal{A} + 1 = \alpha$ & $\mathcal{B} + 1 = \beta$ ▷ The high confidence nodes do not change any more.

9:     $\Pi \leftarrow$ Hard matching obtained by discretizing $X = [x_{ij}]$.

10: **else**

11:     $\alpha \leftarrow \mathcal{A} + 1$, $\beta \leftarrow \mathcal{B} + 1$.

12:     Compute $d_u\Big(\mathcal{F}_{\mathcal{V}_{1i}}(\alpha), \mathcal{F}_{\mathcal{V}_{2j}}(\beta)\Big) = \sum_{r=1}^{|\mathcal{A}|+1} |\mathcal{F}_{\mathcal{V}_{1i}}(\alpha_r) - \mathcal{F}_{\mathcal{V}_{2j}}(\beta_r)|$.

13:     Go to step 5.

$$\phi_{xxx}(x_{ij}, x_{mn}, x_{pq}) = x_{ij}x_{mn}x_{pq}d_t(\mathcal{T}_1, \mathcal{T}_2) + (1 - x_{ij}, x_{mn}, x_{pq})(1 - d_t(\mathcal{T}_1, \mathcal{T}_2)$$

$$d_t(\mathcal{T}_1, \mathcal{T}_2) = w_t^1 |area(\mathcal{T}_1) - area(\mathcal{T}_2)| + \sum_{i=1}^{3} w_t^2 |\angle\mathcal{T}_1^i - \angle\mathcal{T}_2^2| \tag{7}$$

d_t measures the difference between the area and the angles of the triangles. The weights w_t^1 and w_t^2 encode the trade off between preserving areas vs. angles.

2.2   Solving for the PSL Matching via MRF Optimization

Since we bootstrap our PSL matching from the high confidence matches in Section 2.3, we restrict our work to the relaxed version of the problem, while having the entropy term discouraging high uncertainty. We explore: (i) tensor power iteration (TPI) [1], and (ii) successive projection (SP)[1] [11] optimization methods (Section 3). Both TPI and SP provide a soft solution considering global constraints $\sum_i x_{ij} \leq 1$ and $\sum_j x_{ij} \leq 1$ to ensure partial matching and to avoid multiple matchings. Note that in Section 3 the results are provided using TPI.

2.3   Self-Learning

As shown in [6], learning the parameters that control the graph matching is important for improving the matching accuracy. The authors in [6] learn the weights w in (1) using gradient descent-based approach. We instead learn an improved objective function by encoding into the unary term new geometric information from the current high confidence matching. In the learning step of our method, we update α and

---

[1] The SP algorithm is applied to the marginalization matrix computed based on the probability of matching the edges and the cliques [11].

Table 1. Comparison between the different methods in terms of the optimization domain, energy terms, and the self-learning (SL) characteristics

Method	Optimizer Soft vs. Hard	$\phi_x^{(1)}$	$\phi_x^{(2)}$	ϕ_{xx}	ϕ_{xxx}	SL	Objective Function
CVPR09 [7]	SP (Soft)	✓	×	✓	×	×	MRF1
CVPR08 [11]	SP (Soft)	✓	×	✓	✓	×	MRF2
ECCV08 [10]	DD (Hard)	✓	×	✓	×	×	MRF1
CVPR10 [12]	DD (Hard)	✓	×	✓	✓	×	MRF2
PAMI11 [1]	TPI (Soft)	✓	×	✓	✓	×	MRF2
Proposed	TPI (Soft)	✓	✓	✓	✓	✓	MRF2EN

$$\text{MRF1} = \text{func}(\phi_x^{(1)}, \phi_{xx})$$
$$\text{MRF2} = \text{func}(\phi_x^{(1)}, \phi_{xx}, \phi_{xxx})$$
$$\text{MRF1EN} = \text{func}(\phi_x^{(1)}, \phi_x^{(2)}, \phi_{xx})$$
$$\text{MRF2EN} = \text{func}(\phi_x^{(1)}, \phi_x^{(2)}, \phi_{xx}, \phi_{xxx})$$

β in (3), which indicate the indices of $\mathcal{F}$ that should be considered in measuring d_u. As shown in Algorithm 1.1, given the current high confidence matching $x(\mathbb{V}_{\mathcal{AB}})$, i.e. $(\mathcal{A}, \mathcal{B}) = \{(i, j) | x_{ij} > \tau\}$, where τ is a confidence-threshold and $|\mathcal{A}| = |\mathcal{B}| = R$ and R is the total number of the high confidence points, α and β in (3) are updated: $\alpha = \mathcal{A} + 1$ and $\beta = \mathcal{B} + 1$. Therefore,

$$d_u\Big(\mathcal{F}_{\mathcal{V}_{1i}}(\alpha), \mathcal{F}_{\mathcal{V}_{2j}}(\beta)\Big) = \sum_{r=1}^{R+1} |\mathcal{F}_{\mathcal{V}_{1i}}(\alpha_r) - \mathcal{F}_{\mathcal{V}_{2j}}(\beta_r)| \tag{8}$$

The α_r-th entry of $\mathcal{F}_{\mathcal{V}_{1k}}$ represents the distance between the vertex $\mathcal{V}_{1k}$ and $\mathcal{V}_{1\alpha}$. In fact, we are effectively diffusing the binary term to the unary term, since this entry in $\mathcal{F}$ is related to the length agreement between the edges. Figure 1 shows examples of the selected high confidence mappings at different iterations.

3   Results

Given a ground truth matching Π^*, and an estimated mapping Π obtained by discretizing the estimated fuzzy solution $X = [x_{ij}]$ (e.g. applying simple thresholding or the Hungarian algorithm [1,4], where $\Pi_{ij} = 1$ is interpreted as a mapping $\Pi(\mathcal{V}_{1i}) = \mathcal{V}_{2j}$), we use the following error measurement to evaluate the quality of the estimated mapping: $\Delta = \sum |\Pi - \Pi^*|/(N_1 N_2)$. We evaluate our method on synthetic data as well as 56 pairs of real images [2]. Note that we identify the PSLs' coordinates on our real data manually and the number of PSLs in our dataset is varied between 3 and 60.

Our synthetic data follows a setup similar to [7]. A cloud of n_c points are generated. The corresponding points in the second set are constructed by perturbing the n_c points. Then, different number of outliers n_o^1 and n_o^2 (representing disappearing and newly appearing PSLs) are added to the two sets.

In Table 1, we analyse our method and five state of the art point matching algorithms in terms of different characteristics. In summary, CVPR09 [7], CVPR08 [11], ECCV08 [10], CVPR10 [12], PAMI11 [1], and our method, can be implemented by setting the objective function in the form of MRF1, MRF2, or MRF2EN mentioned in Table 1, and applying different optimization approaches. For example, we can arrive to PAMI11 [1] by setting the objective function to MRF2 and using the TPI optimizer. To study

the effectiveness of the entropy term (5), we compare the matching errors resulting
from using a given function, with and without the entropy term; i.e. compare MRF1
vs. MRF1EN and MRF2 vs. MRF2EN. The results in Figure 3 indicate that adding the
entropy term can lead to lower error. The effect of applying different iterations of the
self-learning procedure is shown in Figure 4. The results confirm the usefulness of our
unsupervised learning from high confidence matches. Note that the errors are gradually
decreasing by increasing the number of the iterations.

A comparison between the point matching methods: CVPR09 [7], CVPR08 [11],
PAMI11 [1], and our method on the real data is shown in Figure 5. Note that all the
methods are fed with the normalized coordinates of the PSLs resulting from apply-
ing the skin back-template proposed in [7]. It can be seen that the lowest error is re-
sulting from MRF2EN+SL, i.e. the results of augmenting MRF2EN with the learning
procedure.
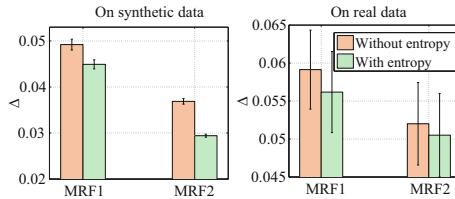
Fig. 3. Usefulness of the entropy term in (5): the y-axis represents the matching error resulting
from applying different objective functions, without (orange bars) and with (green bars) the en-
tropy term
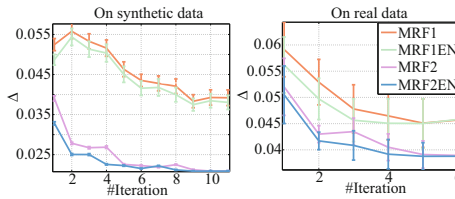
Fig. 4. Learning procedure: the y-axis represents the matching error. Different colors correspond
to different objective functions. The results indicate that iteratively applying the unsupervised
learning procedure (Section 2.3) leads to lower error.
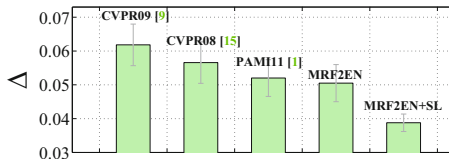
Fig. 5. Comparison between the point matching methods CVPR09 [7], CVPR08 [11], PAMI11
[1], and our method on the real data. Note that MRF2EN+SL has the lowest error.

4 Conclusion

We formulate the PSL matching problem in dermoscopic images as the relaxed labeling of the corresponding association graph in a high order MRF optimization framework. We add a novel entropy term to the objective function encouraging the cost function towards solutions with low uncertainty. We also propose to learn the objective function in a sequential framework by leveraging the high confidence matching of the fuzzy solutions. Although we evaluate the usefulness of the entropy term and the learning procedure on a specific application, the same idea can be used to extend other existing point matching algorithms.

This work can be extended in a number of ways. As mentioned in Section 2.3, for example, the learning step can be generalized for the binary and ternary terms of the matching objective function.

References

1. Duchenne, O., Bach, F., Kweon, I., Ponce, J.: A tensor-based algorithm for high-order graph matching. IEEE TPAMI 1(99), 1–13 (2011) 99, 102, 103, 104
2. Gallagher, R., Rivers, J., Lee, T., Bajdik, C., McLean, D.I., Coldman, A.: Broad-Spectrum Sunscreen Use and the Development of New Nevi in White Children: A Randomized Controlled Trial. JAMA 283(22), 2955–2960 (2000) 103
3. Huang, H., Bergstresser, P.: A new hybrid technique for dermatological image registration. IEEE BIBE, 1163–1167 (2007) 98
4. Kuhn, H.W.: The Hungarian method for the assignment problem. Naval Research Logistic Quarterly 2, 83–97 (1955) 103
5. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: IEEE ICCV, vol. 2, pp. 1482–1489 (2005)
6. Leordeanu, M., Zanfir, A., Sminchisescu, C.: Semi-supervised Learning and Optimization for Hypergraph Matching. In: IEEE ICCV (2011) 102
7. Mirzaalian, H., Hamarneh, G., Lee, T.: Graph-based approach to skin mole matching incorporating template-normalized coordinates. In: IEEE CVPR, pp. 2152–2159 (2009) 98, 99, 101, 103, 104
8. Perednia, D.A., White, R.G.: Automatic registration of multiple skin lesions by use of point pattern matching. CMIG 16(3), 205–216 (1992) 98
9. Roning, J., Riech, M.: Registration of nevi in successive skin images for early detection of melanoma. In: IEEE ICPR, vol. 1, pp. 352–357 (1998) 99
10. Torresani, L., Kolmogorov, V., Rother, C.: Feature Correspondence Via Graph Matching: Models and Global Optimization. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 596–609. Springer, Heidelberg (2008) 99, 100, 103
11. Zass, R., Shashua, A.: Probabilistic graph and hypergraph matching. In: IEEE CVPR, pp. 1–8 (2008) 99, 102, 103, 104
12. Zeng, Y., Wang, C., Yang, W., Gu, X., Samaras, D., Paragios, N.: Dense non-rigid surface registration using high-order graph matching. In: IEEE CVPR, pp. 382–389 (2010) 99, 103

Automatic Categorization of Anatomical Landmark-Local Appearances Based on Diffeomorphic Demons and Spectral Clustering for Constructing Detector Ensembles

Shouhei Hanaoka[1], Yoshitaka Masutani[1,2], Mitsutaka Nemoto[1], Yukihiro Nomura[1], Takeharu Yoshikawa[3], Naoto Hayashi[3], and Kuni Ohtomo[1,2]

[1] Department of Radiology,
[2] Division of Radiology and Biomedical Engineering, Graduate School of Medicine
[3] Department of Computational Diagnostic Radiology and Preventive Medicine,
The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
`hanaoka-tky@umin.ac.jp`

Abstract. A method for categorizing landmark-local appearances extracted from computed tomography (CT) datasets is presented. Anatomical landmarks in the human body inevitably have inter-individual variations that cause difficulty in automatic landmark detection processes. The goal of this study is to categorize subjects (i.e., training datasets) according to local shape variations of such a landmark so that each subgroup has less shape variation and thus the machine learning of each landmark detector is much easier. The similarity between each subject pair is measured based on the non-rigid registration result between them. These similarities are used by the spectral clustering process. After the clustering, all training datasets in each cluster, as well as synthesized intermediate images calculated from all subject-pairs in the cluster, are used to train the corresponding subgroup detector. All of these trained detectors compose a detector ensemble to detect the target landmark. Evaluation with clinical CT datasets showed great improvement in the detection performance.

Keywords: anatomical landmark, diffeomorphic demons, spectral clustering.

1 Introduction

Detection algorithms for anatomical landmark positions have a wide range of applications in medical image processing. For example, algorithms are frequently used to determine the initial condition of statistical shape models for the segmentation of various organs. A simple template matching technique has been frequently used in practice on the assumption that a landmark has sufficiently small inter-individual variations in its local appearance (i.e., the intensities of the voxels around the landmark point). However, many practically important anatomical landmarks in the human body, such as tips of bone structures or bifurcations of vessels, have considerably large inter-individual variations in appearance. Although difficult, detection of these

landmarks is a critical task in medical image processing, such as the precise segmentation of organs [1] or the identification of anatomical anomalies [2]. Therefore, a more reliable method for detecting such landmarks is desired.

It is especially challenging to develop a general but reliable method to detect diverse types of landmarks. One problem is that the intensity of voxels within the landmark-local appearance can greatly change according to the various shapes of the anatomical structure on which the landmark is defined. For example, the inter-individual variety in the length of the 12th rib tip significantly changes the voxel values of the local appearance (Fig. 1). To overcome this problem, a method that can interpolate the shapes—not appearances—between two given training datasets by synthetically augmenting the amount of the training datasets would be beneficial. Developing a shape interpolator for this purpose is one of the goals of this study.
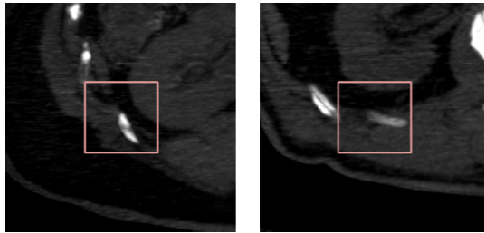
Fig. 1. An example of a landmark difficult to detect due to inter-individual variety (the tip of the 12th right rib in axial images from two datasets)

Another problem is that the shape of the anatomical structure on which a landmark is defined can be multimodal, which means, the structure has several subgroups of shape varieties. If the distribution of appearance has several local maxima (i.e., the distribution cannot be represented by a concave probability function), a significant problem arises if one tries to approximate it by a multidimensional Gaussian distribution (or any other concave distribution). Unfortunately, in our experience in defining and detecting landmarks, this problem is not uncommon (as shown in Fig. 3). In such a case, standard modeling with principal component analysis (PCA) is insufficient. However, at least two different approaches to solve this problem can be considered: 1) sub-categorizing the appearances of training datasets and dividing them into several subsets, or 2) applying machine learning techniques that do not need any assumption of concavity in the probability distribution. The second goal of this study is to develop a method to sub-categorize the appearances (the first of the two approaches).

Up to now, a number of detection methods for multiple landmarks with high reliability have been reported. Seifert et al. [1] reported a framework for detecting landmarks that includes 1) key slice detectors, 2) landmark point detectors that utilize a probabilistic boosting tree with 3D Haar-like features, and 3) a belief propagation algorithm on a predefined inter-landmark graph network. In a previous paper [3], we reported another framework that consists of 1) single-landmark detectors with PCA-based appearance models, 2) false positive (FP) reduction with MadaBoost-based classifiers, and 3) final determination using prior knowledge on inter-landmark distances. Both of these frameworks use a boosting-based approach to handle a wide variety of landmark-local

appearances. In our previous approach, however, it was assumed that the multimodal distribution problem described above affected the framework, because the framework relies on a PCA-based appearance model in the first step.

In this study, we propose a novel method to 1) categorize the given subjects, i.e., training CT datasets, according to landmark-local shape variations of the anatomical structure, 2) synthesize an appearance (a local volume around the landmark) that has the intermediate anatomical shape between any given pair of training datasets, and 3) utilize them to improve the detection performance of landmark detectors through composing a "detector ensemble," which consists of detectors, one of which is designed for only a single subgroup of the local appearance of the target landmark. We used the detector proposed in our previous work [3] as a baseline and improved its performance by replacing the detector with a detector ensemble.

2 Methods

The proposed method consists of three components: 1) shape interpolation and similarity evaluation using the diffeomorphic demons registration algorithm [4], 2) clustering of the appearance variation by using the spectral clustering algorithm [5] and 3) training of subgroup-specific detectors (Fig. 2).

2.1 Shape Interpolation and Similarity Evaluation

2.1.1 Diffeomorphic Demons Registration between Each Appearance Pair

The diffeomorphic demons algorithm is a non-rigid registration algorithm that ensures diffeomorphism of the resulting deformation field [4]. We chose this algorithm because of its invertibility and relatively low computational cost. Invertibility is critical in our application for synthesizing the intermediate shape image.

Let $I_a(\mathbf{p})$, $I_b(\mathbf{p})$ be image functions of the two images to be registered. The vector $\mathbf{p}$ indicates an arbitrary point in the image. Suppose that a 3-D transformation s is defined by a displacement vector field $\mathbf{s}(\cdot)$ such that point $\mathbf{p} + \mathbf{s}(\mathbf{p})$ is warped to point $\mathbf{p}$. Then an image $I_a \circ s$, which is the deformed image of I_a by the transformation s, is defined as

$$(I_a \circ s)(\mathbf{p}) = I_a(\mathbf{p} + \mathbf{s}(\mathbf{p})). \tag{1}$$

We modified the diffeomorphic demons algorithm to be symmetrical for our purpose. (Vercauteren et al. introduced another solution using the Baker-Campbell-Hausdorff (BCH) approximation [4], but we chose this modification because of its simplicity.) The algorithm is as follows:

Algorithm 1 (Diffeomorphic demons with symmetric forces)

1. Initialize the speed image $\mathbf{u}(\mathbf{p})$ to be $(0,0,0)^T$ at any point $\mathbf{p}$ in the image.
2. Calculate a pair of the symmetrical deformation field c_a, c_b by calculating the exponential (in the sense of Lie algebra theory, as described in [4]) of the speed image as follows:

$$c_a = \exp(\mathbf{u}), c_b = \exp(-\mathbf{u}) \tag{2}$$

3. For diffusion-like regularization, apply Gaussian smoothing with kernel size σ_{diff} to the vector fields $\mathbf{c}_a$ and $\mathbf{c}_b$. Let the smoothed results be $\mathbf{s}_a$ and $\mathbf{s}_b$.
4. Update the speed image by

$$\mathbf{u} \to \mathbf{u} - \frac{I_a \circ s_a - I_b \circ s_b}{\|\mathbf{J_p}\|^2 + |I_a \circ c_a - I_b \circ c_b|^2} \cdot \mathbf{J_p}^T \tag{3}$$

Here, $(\mathbf{p})$ is omitted and $\mathbf{J_p} = -\frac{1}{2}\left(\nabla_p^T(I_a \circ s_a) + \nabla_p^T(I_b \circ s_b)\right)$ is a symmetrized Jacobian of the current images.

5. For fluid-like regularization, apply Gaussian smoothing with kernel size σ_{fluid} to the speed vector fields $\mathbf{u}$.
6. Iterate steps 2 through 5 for a sufficient number of times.

This is a fully symmetric version of diffeomorphic demons which deforms each of two images simultaneously so as to fit to the other. Note that the deformation c_a is an inverted version of c_b, and *vice versa*, in the sense that $c_a \circ c_b = c_b \circ c_a = \text{Id}$ (the identity transformation). Finally, the intermediate image is calculated as a mean of two deformed images, or, $\frac{1}{2}\left((I_a \circ s_a)(\mathbf{p}) + (I_b \circ s_b)(\mathbf{p})\right)$.

2.1.2   Calculating the Similarity Measure between Each Appearance Pair

Prior to clustering the landmark-local appearances extracted from the given training datasets, the shape similarity measure must be calculated adequately. In this study, the shape similarity is estimated by the squared norm of the divergence of the deformation vector fields $\mathbf{s}_a$ and $\mathbf{s}_b$ calculated above. That is,

$$\text{Sim}(I_a, I_b) = \exp\left\{-\frac{1}{2\,\sigma_{sim}^2}\int_\Omega \frac{1}{2}\cdot(\|\nabla \cdot \mathbf{s}_a\|^2 + \|\nabla \cdot \mathbf{s}_b\|^2)dv\right\}, \tag{4}$$

where Ω is the domain of the landmark-local appearance and dv is the volume element. The parameter σ_{sim} was determined empirically in this study (the experimental data is omitted), as well as σ_{fluid} and σ_{diff} in 2.1.1.

2.2   Clustering the Landmark-Local Appearances

Spectral clustering methods are algorithms that cluster data samples using eigenvectors of matrices derived from the data. In [5], Ng et al. presented a simple spectral clustering algorithm and the theoretical background. The following is a brief description of the algorithm to cluster n given datasets into k clusters:

Algorithm 2 (Spectral clustering)

1. Form an affinity matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ as its factor A_{ij} equals to the similarity measure between the i-th and j-th appearances. The diagonal elements A_{ii} are set as zero.

2. Define a diagonal matrix $\mathbf{D}$ so that $D_{ii} = \sum_{j=1}^{n} A_{ij}$, and a matrix $\mathbf{L} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$.

3. Find the k largest eigenvectors of $\mathbf{L}$. Let them be $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. Form a matrix $\mathbf{X} = (\mathbf{x}_1 \, \mathbf{x}_2 \dots \mathbf{x}_k) \in \mathbb{R}^{n \times k}$.

4. Form the matrix $\mathbf{Y}$ from $\mathbf{X}$ by renormalizing each row of $\mathbf{X}$, i.e., $Y_{ij} = X_{i,j} / \sqrt{\sum_{j=1}^{n} X_{i,j}^{2}}$.

5. Treat each row of $\mathbf{Y}$ as a sample point in $\mathbb{R}^k$, then cluster these sample points into k clusters via the K-means algorithm.

Here, the number of cluster k is a parameter determined in advance. In this study, we chose k from 2 to 6 for each landmark by the automatic cluster number determination method proposed by Zelnik-Manor et al [6].
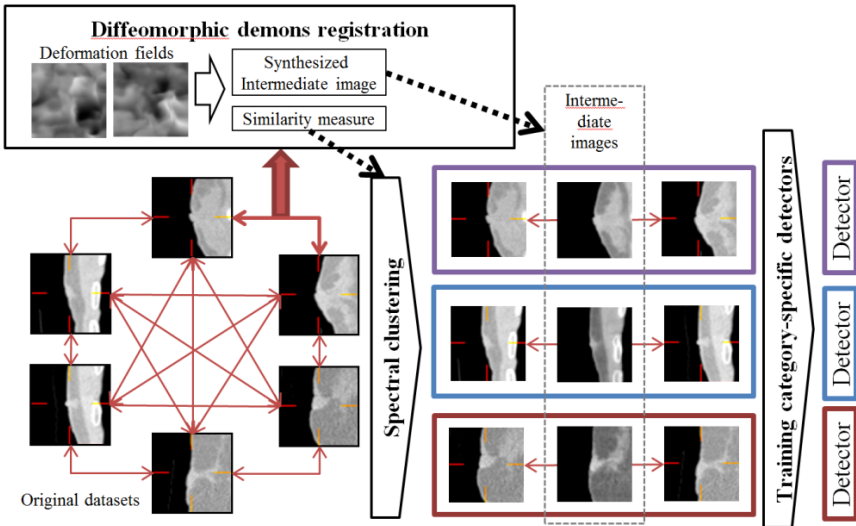
Fig. 2. A diagram of non-rigid registration, clustering and training of category-specific detectors. The images are sagittal cross-sections of the target landmark (left nipple).

2.3   Training of Baseline Detectors and Detector Ensembles

In this study, 50 training datasets were used to train the detector for each landmark. The positions of landmarks were input by a board-certificated radiologist.

The baseline detector. As previously noted, we used the landmark detection framework presented in [3] for evaluation of the proposed method. A detector is composed of two processes, A) PCA-based candidate detection and B) MadaBoost-based likelihood estimation. The detector outputs a list of candidate points with their estimated likelihoods. More details are available in [3]. Note that the proposed method is not

specific to our detector but can be applied to any detection system designed to output a list of candidates with likelihoods, because our algorithm only does clustering and augmenting of the training datasets prior to detector training.

The detector ensemble. After the local appearances of each target landmark from all training datasets were divided into clusters (described in 2.2.), each subgroup (=cluster) detector was trained in the same manner as that of the baseline detector. In the training for candidate detection, not only the original datasets but also the synthesized intermediate images in the cluster were used as the training datasets.

2.4   Experimental Settings

The proposed method was evaluated with 13 landmarks: inferior tip of the sternum, tip of the coccyx, tips of the left/right 12th ribs, umbilicus, superior tips of the right/left kidneys, first bifurcation of the intrahepatic portal vein, inferior tip of the liver, roots of the celiac and superior mesenteric arteries, and right/left nipples. Most of the landmarks were difficult to detect by the baseline detector.

The evaluation was performed with 10 CT datasets (other than the training datasets) without intravenous contrast agent administration. The voxel size of all datasets was $0.977 \times 0.977 \times 1.250$ mm.

In the real landmark detection process, the outputs of the subgroup detectors within any ensemble were integrated as follows. 1) All subgroup detectors, as well as the baseline detector, were applied individually. 2) The resulting candidate lists were concatenated and sorted by likelihood. 3) Finally, only the 100 candidates with the largest likelihoods were used and the others were discarded. For comparison, the outputs of each baseline detector were also truncated to have 100 candidates.

The performance of each detector (baseline or ensemble) was evaluated by a receiver operating characteristic (ROC) analysis by changing the cut-off threshold of the candidate likelihood. The performance was scored by the criterion of the area under the curve (AUC). Each detection result was judged as correct if the detected point was within the range of 2 cm from the manually inputted gold standard point.

3   Experimental Results and Discussions

The results are shown in Table 1. Examples of the clustering results and the ROC curves are shown in Figs. 3 and 4. The AUCs were improved in 11 landmarks (84.6%). The average and standard deviation of the improvements was 0.13 ± 0.12. Remarkable improvements were seen in landmarks even when corresponding baseline detectors showed very poor performances. Therefore, we believe that our method successfully clustered and augmented the training datasets with multimodal variation.

In a related study, Heckemann et al. [7] proposed an automatic segmentation method by non-rigidly registering multiple labeled images (atlases) into the given unseen

image and integrating the results to form the final result. This method showed excellent results and seems to be effective for handling multimodal variation. However, compared to their multi-atlas approach, our approach has the benefit of less computational cost in processing unseen images, in return for a larger cost of registration in the training process. Therefore, we believe that our approach can be an alternative choice for handling the multimodal inter-individual variety of organ shapes.
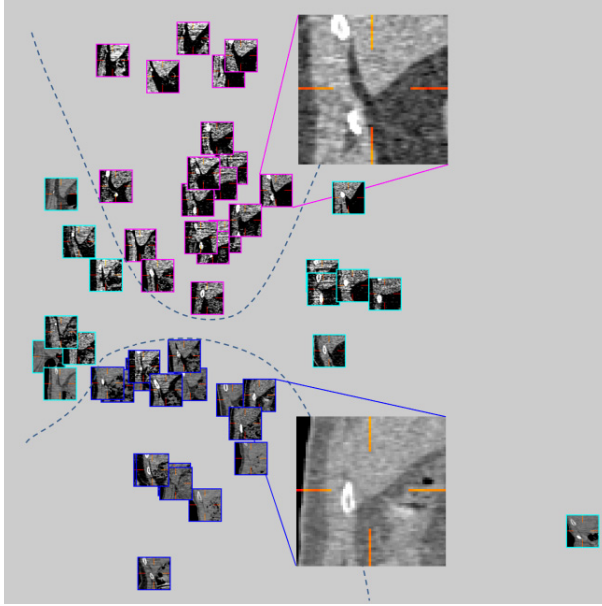
Fig. 3. An example of landmark appearance variation (inferior tip of the liver). Shown are the coronal cross-sections, mapped according to the dimension reduction result by a Laplacian eigenmap. The curves indicate our spectral clustering result. Note that appearances in the upper subgroup have an obvious fat pad between the liver and the abdominal wall muscles, while those in the lower subgroup have only a trace amount of fat pad.
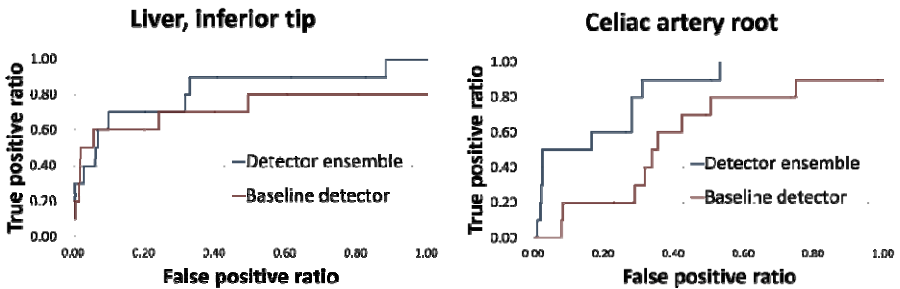
Fig. 4. Two examples of ROC curves (inferior tip of the liver and root of the celiac artery)

Table 1. The AUC (area under the curve) results of the baseline detector and the detector classifier

LM name	Sternum, inf. tip	Tip of coccyx	Tip of rt. 12th rib	Tip of lt. 12th rib	Umbilicus	Rt. Kidney, sup. tip	Lt. Kidney, sup. tip	Bifurcation of IHPV	Liver, inf. tip	Celiac artery root	SMA root	Rt. nipple	Lt. nipple	Average
Baseline	0.31	1.00	0.80	0.61	0.75	0.80	0.66	0.57	0.72	0.59	0.66	0.31	0.07	0.60
Ensemble	0.46	1.00	0.78	0.66	0.82	0.77	0.69	0.70	0.82	0.84	0.81	0.60	0.48	0.73
Improve-ment	0.15	0.00	-0.02	0.05	0.07	-0.03	0.03	0.13	0.10	0.25	0.15	0.29	0.41	0.13

4   Conclusion

A method for clustering and augmenting landmark-local appearances prior to training of anatomical landmark detectors was presented. Ensembles of cluster-specific detectors showed large improvements in detection performance, even when detection by the baseline detector was significantly difficult.

References

1. Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body CT data. In: Samei, E., Hsieh, J. (eds.) Medical Imaging 2009: Physics of Medical Imaging. Proceedings of the SPIE, vol. 7258, pp. 725902–725902-8 (2009)
2. Hanaoka, S., Masutani, Y., Nemoto, M., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: Probabilistic Modeling of Landmark Distances and Structure for Anomaly-proof Landmark Detection. In: Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy, pp. 159–169 (2011)
3. Nemoto, M., Masutani, Y., Hanaoka, S., Nomura, Y., Yoshikawa, T., Hayashi, N., Yoshioka, N., Ohtomo, K.: A unified framework for concurrent detection of anatomical landmarks for medical image understanding. In: Proc. SPIE, vol. 7962, p. 79623E (2011)
4. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage 45(1) suppl. 1, S61–S72 (2009)
5. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 14, pp. 849–856 (2001)
6. Zelnik-Manor, L., Perona, P.: Self-Tuning Spectral Clustering. In: Advances in Neural Information Processing Systems, vol. 17, pp. 1601–1608 (2004)
7. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33(1), 115–126 (2006)

A Novel Approach for Global Lung Registration Using 3D Markov-Gibbs Appearance Model

Ayman El-Baz[1], Fahmi Khalifa[1], Ahmed Elnakib[1], Matthew Nitzken[1],
Ahmed Soliman[1], Patrick McClure[1], Mohamed Abou El-Ghar[2],
and Georgy Gimel'farb[3]

[1] BioImaging Laboratory, Bioengineering Department,
University of Louisville, Louisville, KY, USA
[2] Radiology Department, Urology and Nephrology Center,
University of Mansoura, Mansoura 35516, Egypt
[3] Department of Computer Science, University of Auckland, Auckland, New Zealand

Abstract. A new approach to align 3D CT data of a segmented lung object with a given prototype (reference lung object) using an affine transformation is proposed. Visual appearance of the lung from CT images, after equalizing their signals, is modeled with a new 3D Markov-Gibbs random field (MGRF) with pairwise interaction model. Similarity to the prototype is measured by a Gibbs energy of signal co-occurrences in a characteristic subset of voxel pairs derived automatically from the prototype. An object is aligned by an affine transformation maximizing the similarity by using an automatic initialization followed by a gradient search. Experiments confirm that our approach aligns complex objects better than popular conventional algorithms.

1 Introduction

Image registration aligns two or more images of similar objects taken at different times, from different viewpoints, and/or by different sensors. The images are geometrically transformed to ensure their close similarity. Registration is a crucial step in many applied image analysis tasks, e.g., early diagnosis of detected lung nodules. One of the most compelling motivations for identifying a potentially malignant nodule is to assess its growth rate. To quantify the growth rate of a nodule, one must be able to measure the volume of nodules and identify corresponding nodules in a follow-up scans. The principal difficulty in estimating the nodule growth rate is automatic identification and registration (alignment) of the corresponding nodules in follow-up scans. Registration of the lung tissues is a challenging task due to large displacements between them in successive CT scans, which may be caused by variation in respiratory volumes and patient positioning. For these reasons, the registration of the successive CT lung data taken at different times is the main goal of this paper.

Most of the known registration methods fall into two main categories: feature-based and area-based techniques [1]. Feature-based techniques use sparse geometric features such as points, curves, and/or surface patches, and their

correspondences to compute an optimal transformation. Area-based methods, such as the classical least square correlation, directly match image signals to avoid feature extraction [2]. More powerful mutual information (MI) based image registration [3] exploits a probabilistic similarity measure that allows for more general types of signal deviations than correlation. The statistical dependency between two data sets is measured by comparing a joint empirical distribution of the corresponding signals in the two images to the joint distribution of the independent signals (e.g., see [4,5] for more details about the existing approaches for medical images registration).

In this paper, we consider a more general case of registering 3D segmented lung data to a prototype (the reference data, usually the CT collected at the first visit of the patient) with similar visual appearance under their relative 3D affine transformations and monotone variations of signal correspondences. To suppress the variations between the images, all test and training images were equalized to the same signal ranges, thus having the same dimensions in their co-occurrence matrices. Generally, any equalization scheme can be used. The co-registered equalized images are described with a characteristic subset of signal co-occurrence statistics. The description implicitly "homogenizes" the images, i.e., considers them as spatially homogeneous patterns with the same statistics. Our approach differs from the feature-based registration in that the statistics characterize the whole object, and from conventional area-based techniques in that the similarities between the statistics rather than pixel-to-pixel correspondences are measured.

2   MGRF Based Image Registration

2.1   Basic Notation

- $\mathcal{Q} = \{0, \ldots, Q - 1\}$ – a finite set of scalar image signals (e.g., gray levels).
- $\mathbf{R} = [(x, y, z) : x = 0, \ldots, X - 1; y = 0, \ldots, Y - 1; z = 0, \ldots, Z - 1]$ – a 3D arithmetic lattice supporting digital low dose CT (LDCT) image data $g : \mathbf{R} \to \mathcal{Q}$.
- $\mathbf{R_p} \subset \mathbf{R}$ – an arbitrary-shaped part of the lattice occupied by a prototype.
- $\mathcal{N} = \{(\xi_1, \eta_1, \zeta_1), \ldots, (\xi_n, \eta_n, \zeta_n)\}$ – a finite set of (x, y, z)-coordinate offsets defining neighboring voxels, or neighbors $\{((x + \xi, y + \eta, z + \zeta), (x - \xi, y - \eta, z - \zeta)) : (\xi, \eta, \zeta) \in \mathcal{N}\} \wedge \mathbf{R_p}$ interacting with each voxel $(x, y, z) \in \mathbf{R_p}$.
- T – an indicator of vector or matrix transposition.

The set $\mathcal{N}$ yields a 3D neighborhood graph on $\mathbf{R_p}$ describing translation invariant pairwise interactions between the voxels with $|\mathcal{N}|$ families $\mathcal{C}_{\xi,\eta,\zeta}$ of the 2nd-order cliques $c_{\xi,\eta,\zeta}(x, y, z) = ((x, y, z), (x + \xi, y + \eta, z + \zeta))$ shown in Fig. 1.

Quantitative interaction strengths for the clique families are given by a vector $\mathbf{V}^{\mathsf{T}} = \left[\mathbf{V}_{\xi,\eta,\zeta}^{\mathsf{T}} : (\xi, \eta, \zeta) \in \mathcal{N}\right]$ of potentials $\mathbf{V}_{\xi,\eta,\zeta}^{\mathsf{T}} = \left[V_{\xi,\eta,\zeta}(q, q') : (q, q') \in \mathcal{Q}^2\right]$ being functions of signal co-occurrences in the cliques.

Data Normalization: To account for possible monotone (order-preserving) changes of signals, e.g., due to different sensor characteristics, every LDCT data set is equalized using the cumulative probability distribution of its signals.

Markov-Gibbs Random Field-Based Appearance Model: The main idea of learning the appearance model using an MGRF is to find the neighborhood system (i.e., the subsets of voxels which have strong relations with each current voxel) and to estimate the interaction between each two voxels in this neighborhood system (see Fig. 1).
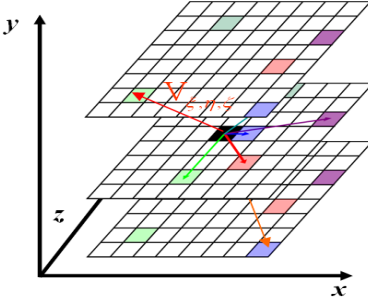
Fig. 1. Pairwise voxel interaction system
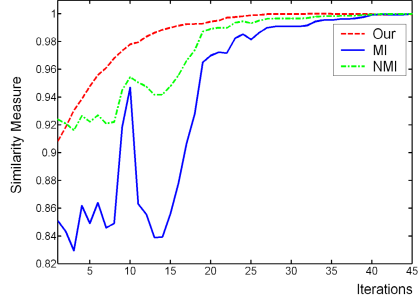
Fig. 2. Values of the Gibbs energy, MI, and NMI at the successive steps of the gradient ascent based search

In a generic MGRF with multiple pairwise interaction in Fig. 1 [6], the Gibbs probability $P(g) \propto \exp(E(g))$ of an object g aligned with the prototype g° on $\mathbf{R_p}$ is specified with the Gibbs energy $E(g) = |\mathbf{R_p}|\mathbf{V}^\mathsf{T}\mathbf{F}(g)$. Here,

- $\mathbf{F}^\mathsf{T}(g)$ is the vector of scaled empirical probability distributions of signal co-occurrences over each clique family: $\mathbf{F}^\mathsf{T}(g) = [\rho_{\xi,\eta,\zeta}\mathbf{F}_{\xi,\eta,\zeta}^\mathsf{T}(g) : (\xi,\eta,\zeta) \in \mathcal{N}]$;
- $\rho_{\xi,\eta,\zeta} = \frac{|\mathcal{C}_{\xi,\eta,\zeta}|}{|\mathbf{R_p}|}$ is the relative size of the family $\mathcal{C}_{\xi,\eta,\zeta}$, and
- $\mathbf{F}_{\xi,\eta,\zeta}(g)$ is the vector of empirical probabilities for this family: $\mathbf{F}_{\xi,\eta,\zeta}(g) = [f_{\xi,\eta,\zeta}(q,q'|g) : (q,q') \in \mathcal{Q}^2]^\mathsf{T}$ where
  - $f_{\xi,\eta,\zeta}(q,q'|g) = \frac{|\mathcal{C}_{\xi,\eta,\zeta;q,q'}(g)|}{|\mathcal{C}_{\xi,\eta,\zeta}|}$ are empirical signal co-occurrence probabilities, and
  - $\mathcal{C}_{\xi,\eta,\zeta;q,q'}(g) \subseteq \mathcal{C}_{\xi,\eta,\zeta}$ is a subfamily of the cliques $c_{\xi,\eta,\zeta}(x,y,z)$ supporting the co-occurrence ($g_{x,y,z} = q$, $g_{x+\xi,y+\eta,z+\zeta} = q'$) in g.

The co-occurrence distributions and the Gibbs energy for the object are determined over $\mathbf{R_p}$, i.e., within the prototype boundary after an object is affinely aligned with the prototype. To account for the affine transformation, the initial image is resampled to the back-projected $\mathbf{R_p}$ by bilinear interpolation.

The appearance model consists of the neighborhood $\mathcal{N}$ and the potential $\mathbf{V}$ to be learned from the prototype. Below we will show how to estimate $\mathcal{N}$ and $\mathbf{V}$ for the lung tissues from the LDCT images.

Learning the Potentials: In the first approximation, the maximum likelihood estimate (MLE) of $\mathbf{V}$ is proportional to the scaled and centered empirical co-occurrence distributions for the prototype[1]:

$$\mathbf{V}_{\xi,\eta,\zeta} = \lambda \rho_{\xi,\eta,\zeta} \left(\mathbf{F}_{\xi,\eta,\zeta}(g^\circ) - \frac{1}{Q^2}\mathbf{U} \right); \ (\xi,\eta,\zeta) \in \mathcal{N}$$

where $\mathbf{U}$ is the vector with unit components. The common scaling factor λ is also computed analytically; it is approximately equal to Q^2 if $Q \gg 1$ and $\rho_{\xi,\eta,\zeta} \approx 1$ for all $(\xi,\eta,\zeta) \in \mathcal{N}$. In our case it can be set to $\lambda = 1$ because the registration uses only relative potential values and energies.

Learning the Characteristic Neighbors: To find the characteristic neighborhood set $\mathcal{N}$, the relative energies $E_{\xi,\eta,\zeta}(g^\circ) = \rho_{\xi,\eta,\zeta} \mathbf{V}_{\xi,\eta,\zeta}^\mathsf{T} \mathbf{F}_{\xi,\eta,\zeta}(g^\circ)$, i.e., the scaled variances of the corresponding empirical co-occurrence distributions for the clique families, are compared for a large number of possible candidates.
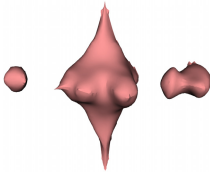
Fig. 3. The 3D neighborhood system estimated for the lung tissues

Table 1. Alignment errors (in mm) for the expert-identified landmarks over the 100 data sets, and the execution times comparing the proposed approach against MI- and NMI-based methods. SD stands for standard deviation.

	Our	MI	NMI
Mean	1.9	5.1	4.8
SD	1.1	2.3	1.9
Time (min)	12	9.0	9.0
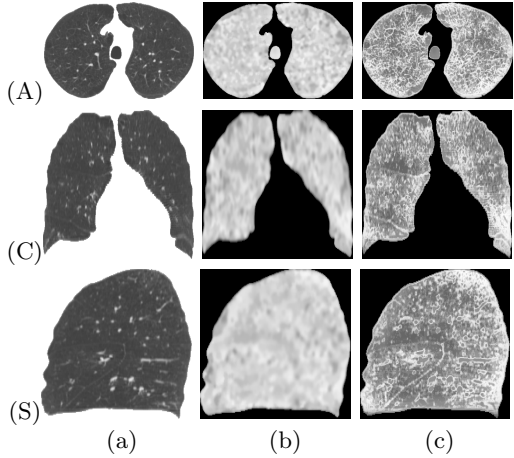
Fig. 4. 3D voxel-wise Gibbs energies projected onto the 2D axial (A), coronal (C), and sagittal (S) planes for visualization: 2D slices of the original LDCT images (a) and the voxel-wise Gibbs energies for $|\mathcal{N}| = 275$ which is estimated using LCDG model (b) and $|\mathcal{N}| = 3927$ which is estimated using tradition EM algorithm (c)

To automatically select the characteristic neighbors, we consider an empirical probability distribution of the energies as a mixture of a large "non-characteristic" low-energy component and a considerably smaller characteristic high-energy component: $P(E) = \pi P_{\mathrm{lo}}(E) + (1-\pi)P_{\mathrm{hi}}(E)$. The components $P_{\mathrm{lo}}(E)$ and $P_{\mathrm{hi}}(E)$ are

---

[1] For complete proof, please see: https://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/supplemental-materials

of arbitrary shape and thus are approximated with linear combinations of positive and negative Gaussians (LCDG) using efficient Expectation-Maximization-based algorithms introduced in [7] (the latter estimate both the components and the prior π). The intersection of the estimated mixture components using the LCDG model gives an energy threshold θ for selecting the characteristic neighbors: $\mathcal{N} = \{(\xi, \eta) : E_{\xi,\eta}(g^\circ) \geq \theta\}$ where $P_{\text{hi}}(\theta) \geq P_{\text{lo}}(\theta)\pi/(1 - \pi)$. The above example results in the threshold $\theta = 28$ producing 275 characteristic neighbors shown in Fig. 3. Figure 4 presents the relative 3D voxel-wise Gibbs energies $e_{x,y,z}(g^\circ)$ for this system:

$$e_{x,y,z}(g^\circ) = \sum_{(\xi,\eta,\zeta)\in\mathcal{N}} V_{\xi,\eta,\zeta}(g^\circ_{x,y,z}, g^\circ_{x+\xi,y+\eta,z+\zeta})$$

Appearance-Based Registration. The desired affine transformation of an object g corresponds to a local maximum of its relative energy $E(g_\mathbf{a}) = \mathbf{V}^\mathsf{T}\mathbf{F}(g_\mathbf{a})$ under the learned appearance model $[\mathcal{N}, \mathbf{V}]$. Here, $g_\mathbf{a}$ is the part of the object image reduced to $\mathbf{R_p}$ by a 3D affine transformation $\mathbf{a} = [a_{11}, \ldots, a_{34}]$: $x' = a_{11}x + a_{12}y + a_{13}z + a_{14}$; $y' = a_{21}x + a_{22}y + a_{23}z + a_{24}$; $z' = a_{31}x + a_{32}y + a_{33}z + a_{34}$. Its initial step is a pure translation ($a_{11} = a_{22} = a_{33} = 1$; $a_{12} = a_{13} = a_{21} = a_{23} = a_{31} = a_{32} = 0$) ensuring the most "energetic" overlap between the object and prototype. In other words, the chosen initial position ($a^*_{14}, a^*_{24}, a^*_{34}$) in Fig. 5 maximizes the energy. Then, the gradient ascent based search for the local energy maximum closest to the initialization selects all the 12 parameters $\mathbf{a}$. Note that this gradient-based optimization was used for all tested registration methods to estimate the goal parameters.
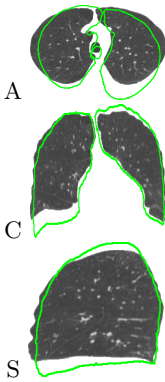
Fig. 5. Initialization of the proposed global registration algorithm
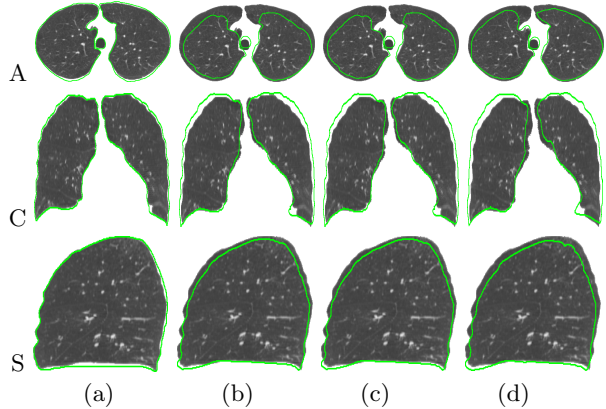
Fig. 6. Global registration results: our (a), MI-based (b), NMI-based (c), and SIFT-based (d) algorithms. These registration results were obtained from the segmented lung data without any pre-processing steps.

3   Experimental Results

The proposed registration approach has been tested on the clinical data sets that have been collected from 100 patients at the end-expiration breath hold (EE-BH). Each patient has two LDCT scans, with the three-month period between each two successive scans. This clinical database was collected by the LDCT scan protocol using a multidetector GE Light Speed Plus scanner (General Electric, Milwuakee, USA) with the following scanning parameters: the slice thickness of 2.5 mm reconstructed every 1.5 mm; scanning pitch 1.5 mm; 140 KV; 100 MA; and the field-of-view 36 cm.

Results of the proposed global alignment of two lungs are shown in Figs. 6 and 7. It is clear from Fig. 6 (a) and Fig. 7 (c,d) that there are small misalignment at the lung surface, this is due to local deformation of lung tissues which come from breathing and heart beats. This can be handle by applying any local deformation transformation model (e.g., 3D cubic splines as a local transformation model [8]) as shown in Fig. 7 (f).

| (a) | (b) | (c) | (d) | (e) | (f) |
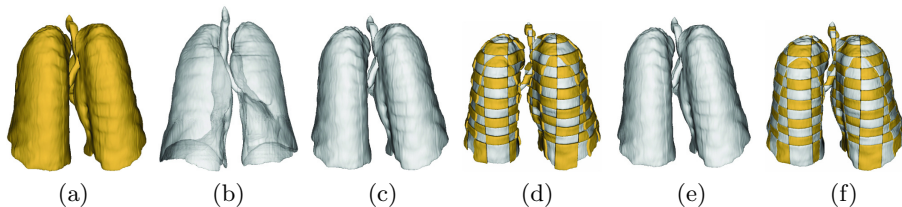
Fig. 7. 3D global and local registration: (a) the reference data, (b) the target data, (c) the target data after a 3D affine transformation, (d) the checkerboard visualization to show the motion of lung tissues, (e) non-rigid registration based using 3D cubic splines as a local transformation model [8], and (f) the checkerboard visualization to show the quality of the proposed local deformation model

To highlight the advantages of the proposed registration approach, we compared, on segmented lung data, our global alignment to three popular conventional techniques, namely, to the area-based registration by mutual information (MI) [3] or normalized MI (NMI) [9] and to the feature-based registration that establishes correspondences between the images with 3D scale-invariant feature transform (SIFT) [10].

To clarify why the MI- or NMI-based alignment is less accurate, Fig. 2 compares the MI/NMI and Gibbs energy values for a sequence of affine parameters that appear at successive steps of the gradient ascent-based search for the maximum similarity in terms of mutual information or energy. Both the MI and NMI have many outstanding local maxima that potentially hinder the search, whereas the energy is much smoother and practically unimodal in these experiments. The 3D SIFT-based alignment fails because it cannot establish accurate correspondences between the similar lung areas.
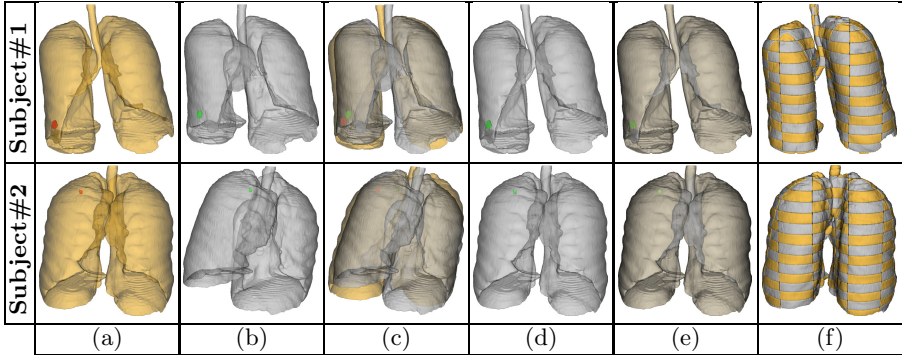
Fig. 8. Registration results of the proposed validation control study: (a) scan #1 before the movement of the patient; (b) scan #2 after patients rotation and translation; (c) the two scans superposed; (d) the registration results of the proposed approach; (e) superposition of the registered scan #2 with the reference scan #1; and (f) checkerboard visualization to show the quality of the proposed registration

Validation: To validate the proposed approach, we have performed a control study using two subjects, each with one solid nodule. The CT data was collected with the same scanning protocol described above. In this new control study, we acquired two scans from the same patient at the EE-BH. The only difference between the two scans was that the patient was asked to make a global rotation and translation after the first scan. The purpose of this control study was to get the data from the same patient with minimal local deformation by breathing in order to test the proposed 3D global registration approach. Figure 8 illustrates the registration results of the proposed approach. We used scan #1 before the patients movement as a reference image to register scan #2 after the patient movement. The superposition of the registered and reference data and the checkerboard visualization in Fig. 8(e,f) demonstrates the high quality of the proposed registration. The average registration error is 1.4 mm with standard deviation ±0.7 mm. This estimated error is based on calculating the Euclidean distance between 250 landmark points manually selected by a radiologist on the reference and the registered target data.

Additionally, our registration approach was validated on the 100 data sets based on using anatomical landmark correspondences selected by a radiologist. After the target data is aligned to the reference data, 10 anatomical landmarks, between both the reference and registered target data, are identified by the radiologist. The registration accuracy is quantitatively assessed by calculating the Euclidian distance for each expert-identified landmark on the registered data and its correspondence on the reference data. The error statistics for this validation experiment as well as the average execution time for our approach and the MI- and NMI-based methods are summarized in Table 1.

4   Conclusions

In this paper we introduced a new approach to align 3D CT data of a lung object with a given prototype whose appearance is modeled with a new 3D Markov-Gibbs random field with pairwise interaction model. Experimental results confirm that lung registration based on our new Markov-Gibbs appearance model is more robust and accurate than popular conventional algorithms. Moreover, it is worth mentioning that the proposed registration approach is not limited only to lung objects, but it is also suitable for registering any 3D texture medical objects. Furthermore, the proposed approach can be integrated with any nonrigid registration algorithm (e.g., cubic B-Splines based techniques). In our future work will use the evaluation framework proposed by van de Kraats et al. [11] to evaluate the performance of our registration approach against the MI- and NMI-based approaches as well as higher-order MI-based techniques (e.g., [12]).

References

1. Zitova, B., Flusser, J.: Image Registration Methods: A Survey. Image Vis. Comput. 21, 977–1000 (2003)
2. Pope, P., Theiler, J.: Automated Image Registration (AIR) of MTI imagery. In: Proc. SPIE 5093, vol. 27, pp. 294–300 (2003)
3. Viola, P.: Alignment by maximization of mutual information. Ph.D. Dissertation, MIT, Cambridge, MA (1995)
4. Pluim, J., Maintz, J., Viergever, M.: Mutual-information Based Registration of Medical Images: A survey. IEEE Trans. Medical Imaging 22(8), 986–1004 (2003)
5. Khalifa, F., Beache, G.M., Gimel'farb, G., Suri, J., El-Baz, A.: State-of-the-Art Medical Images Registration Methodologies: A Survey. In: El-Baz, A., Acharya, R., Mirmedhdi, M., Suri, J. (eds.) Handbook of Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies, vol. 1, ch. 9, pp. 235–277. Springer (2011)
6. Farag, A., El-Baz, A., Gimel'farb, G.: Precise Segmentation of Multimodal Images. IEEE Trans. Image Process. 15(4), 952–968 (2006)
7. El-Baz, A., Elnakib, A., Khalifa, F., Abou El-Ghar, M., McClure, P., Soliman, A., Gimel'farb, G.: Precise Segmentation of 3D Magnetic Resonance Angiography. IEEE Trans. Biomed. Eng. 59(7), 2019–2029 (2012)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid Registration Using Free-Form Deformations: Application to Breast MR Images. IEEE Trans. Medical Imaging 18(8), 712–721 (1999)
9. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition 32(1), 71–86 (1999)
10. Aly, A.: Local Features Invariance Beyond 2D Gray Spaces. Ph.D. Dissertation, University of Louisville, Louisville, KY (2007)
11. Van de Kraats, E.B., Penney, G.P., Tomazevic, D., Van Walsum, T., Niessen, W.J.: Standardized Evaluation Methodology for 2-D-3-D Registration. IEEE Trans. Medical Imaging 24(9), 1177–1189 (2005)
12. Rueckert, D., Clarkson, M.J., Hill, D.L.G., Hawkes, D.J.: Non-rigid Registration using Higher-Order Mutual Information. In: Proc. SPIE 3979, vol. 27, pp. 439–447 (2000)

Analytic Regularization of Uniform Cubic B-spline Deformation Fields

James A. Shackleford[1], Qi Yang,[2] Ana M. Lourenço[3], Nadya Shusharina[1], Nagarajan Kandasamy[4], and Gregory C. Sharp[1]

[1] Department of Radiation Oncology,
Massachusetts General Hospital, Boston, MA 02114, USA
[2] Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Germany
[3] Institute of Biophysics and Biomedical Engineering,
Faculty of Sciences of the University of Lisbon, Lisbon, Portugal
[4] Electrical and Computer Engineering Department,
Drexel University, Philadelphia, PA 19104, USA

Abstract. Image registration is inherently ill-posed, and lacks a unique solution. In the context of medical applications, it is desirable to avoid solutions that describe physically unsound deformations within the patient anatomy. Among the accepted methods of regularizing non-rigid image registration to provide solutions applicable to medical practice is the penalty of thin-plate bending energy. In this paper, we develop an exact, analytic method for computing the bending energy of a three-dimensional B-spline deformation field as a quadratic matrix operation on the spline coefficient values. Results presented on ten thoracic case studies indicate the analytic solution is between 61–1371x faster than a numerical central differencing solution.

Keywords: deformable registration, b-spline, analytic regularization, thin-plate, bending energy

1 Introduction

B-spline based deformable registration has become a popular method for deriving coordinate system transforms between image volumes exhibiting complex local variations due to its compact local support, rapid computation, and applicability to both single and multi-modalities. Such transforms allow non-rigid structures to be mapped between images and provide quantitative measure of local motion and volumetric change over time. Consequently, deformable registration has played an important role in advancing numerous fields of research and applied medicine including Alzheimer's disease [1], schizophrenia [2], generalized brain development [3], image-guided surgery [4,5], image guided radiotherapy [6,7], motion estimation [8] and time-evolution visualization [9].

Due to the inherent ill-posed nature of image registration, the existence of a unique mapping is not guaranteed and the solution space must, therefore, be confined to only physically meaningful transforms. To this end, several regularization methods have been proposed: Rueckert *et al.* propose penalizing high

thin-plate bending energy [10]. Rohlfing *et al.* propose penalizing local devia-
tions from a unity Jacobian determinate [11]. Miller *et al.* propose minimizing
linear elastic energy[12]. Li *et al.* enforce a maximum delta between adjacent B-
spline coefficients [13], whereas Chun and Fessler propose encouraging invertible
diffeomorphic transforms by placing more complex constraints upon coefficients.
This paper builds upon [10] by introducing a fast analytic method for computing
the thin-plate bending energy penalty via a set of static matrix operators.

2   Theory

Here we develop the necessary theory to compute the bend-
ing energy of a three-dimensional deformation field parame-
terized by a uniform cubic B-spline basis. Given a uniformly
spaced control-point grid as shown in Fig. 1, the bending
energy of the entire deformation may be expressed as a lin-
ear combination of the bending energies computed within
the individual regions of the grid. Therefore, our approach
is to develop an operator that computes the bending energy
within a region as a function of the B-spline control points
that support the region.
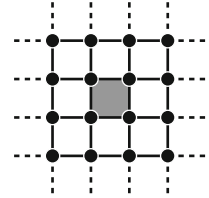
Fig. 1. 2D region
supported by 16
control-points

Given a three dimensional fixed image F with voxel coordinates $\boldsymbol{\theta} = x, y, z$
and voxel intensity $f = F(\boldsymbol{\theta})$ and moving image M with voxel coordinates
$\boldsymbol{\phi} = x_2, y_2, z_2$ and voxel intensity $m = M(\boldsymbol{\phi})$ representing the same underlying
anatomy as F within the image overlap domain $\boldsymbol{\Omega}$, the two images F and M are
said to be registered when cost function

$$C = \sum_{\mathbf{T}(\boldsymbol{\theta}) \in \boldsymbol{\Omega}} \boldsymbol{\Psi}\left(f, m\right) + \lambda S \tag{1}$$

is optimized according to the similarity metric $\boldsymbol{\Psi}$ under the coordinate mapping
$\mathbf{T}(\boldsymbol{\theta}) = \boldsymbol{\theta} + \boldsymbol{\nu}$. Here $\boldsymbol{\nu}$ is the dense vector field defined for every voxel $\boldsymbol{\theta} \in \boldsymbol{\Omega}$,
which is assumed capable of providing a good one-to-one mapping from F to M.
The smoothness S of $\boldsymbol{\nu}$ is added to C with weight λ to drive $\mathbf{T}$ to a physically
meaningful coordinate map. When represented sparsely via the uniform cubic
B-spline basis, the vector field $\boldsymbol{\nu}$ is parameterized by the set of B-spline basis
coefficients $\boldsymbol{P}_{i,j,k} = \boldsymbol{p}_x, \boldsymbol{p}_y, \boldsymbol{p}_z$, where:

$$\boldsymbol{p}_x = \begin{bmatrix} p_{x,0,0,0} \\ \vdots \\ p_{x,I,J,K} \end{bmatrix}, \boldsymbol{p}_y = \begin{bmatrix} p_{y,0,0,0} \\ \vdots \\ p_{y,I,J,K} \end{bmatrix}, \boldsymbol{p}_z = \begin{bmatrix} p_{z,0,0,0} \\ \vdots \\ p_{z,I,J,K} \end{bmatrix} \tag{2}$$

are defined for $n = I \times J \times K$ control-points with real world spacing $\boldsymbol{r} = r_x, r_y, r_z$. From this new basis, the vector field may be expressed at a point $\boldsymbol{\theta}$

with Euclidean coefficients $\boldsymbol{\nu}$ computed via the following tensor product using the 64 B-spline coefficients supporting $\boldsymbol{\theta}$:

$$
\nu_x = \sum_{i=0}^{3}\sum_{j=0}^{3}\sum_{k=0}^{3} \boldsymbol{p}_{i,j,k} \sum_{a=0}^{3} \mathbf{Q}_x^{(\delta_x)}(i,a)\boldsymbol{x}(a) \sum_{b=0}^{3} \mathbf{Q}_y^{(\delta_y)}(j,b)\boldsymbol{y}(b) \sum_{c=0}^{3} \mathbf{Q}_z^{(\delta_z)}(k,c)\boldsymbol{z}(c)
$$

$$(3)$$

for the x-dimension and similarly for the y- and z-dimensions. Here

$$
\boldsymbol{x} = \begin{bmatrix} 1 & x & x^2 & x^3 \end{bmatrix}^{\mathbf{T}}
$$

$$(4)$$

forms a Cartesian basis and $\boldsymbol{y}$ and $\boldsymbol{z}$ are defined similarly. The matrices $\mathbf{Q}_{\mathbf{x}}^{(\delta)}$, $\mathbf{Q}_{\mathbf{y}}^{(\delta)}$, and $\mathbf{Q}_{\mathbf{z}}^{(\delta)}$ are defined by

$$
\mathbf{Q}_x^{(\delta)} = \mathbf{B}\mathbf{R}_x\boldsymbol{\Delta}^{(\delta)} \quad \mathbf{Q}_y^{(\delta)} = \mathbf{B}\mathbf{R}_y\boldsymbol{\Delta}^{(\delta)} \quad \mathbf{Q}_z^{(\delta)} = \mathbf{B}\mathbf{R}_z\boldsymbol{\Delta}^{(\delta)}
$$

$$(5)$$

where $\mathbf{B}$ forms the cubic B-spline basis and $\mathbf{R}_x$, $\mathbf{R}_y$, and $\mathbf{R}_z$ confine the evaluation of the B-spline basis to $\in [0,1]$:

$$
\mathbf{B} = \frac{1}{6}\begin{bmatrix} 1 & -3 & 3 & -1 \\ 4 & 0 & -6 & 3 \\ 1 & 3 & 3 & -3 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{r_x} & 0 & 0 \\ 0 & 0 & \frac{1}{r_x^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r_x^3} \end{bmatrix},
$$

$$(6)$$

The matrix $\boldsymbol{\Delta}^{(\delta)}$ is defined thusly for $\delta \in [0,2]$

$$
\boldsymbol{\Delta}^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{\Delta}^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}, \quad \boldsymbol{\Delta}^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \end{bmatrix}.
$$

$$(7)$$

and provides a convenient method for obtaining $\boldsymbol{\nu}'$ and $\boldsymbol{\nu}''$ w.r.t to the Euclidean basis as required by the calculation of the smoothness penalty [10]:

$$
S = \int_{\boldsymbol{\Omega}} \left(\frac{\partial^2 \boldsymbol{\nu}}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \boldsymbol{\nu}}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \boldsymbol{\nu}}{\partial z^2}\right)^2
$$
$$
+ 2\left[\left(\frac{\partial^2 \boldsymbol{\nu}}{\partial xy}\right)^2 + \left(\frac{\partial^2 \boldsymbol{\nu}}{\partial xz}\right)^2 + \left(\frac{\partial^2 \boldsymbol{\nu}}{\partial yz}\right)^2\right] d\boldsymbol{x}.
$$

$$(8)$$

We may obtain expressions for these derivative terms by referring to (3) and expanding the triple summation over (i,j,k) to the 64×1 vector:

$$
\boldsymbol{\gamma}^{(\delta_x,\delta_y,\delta_z)} =
$$
$$
\begin{bmatrix} \left(\sum_a \mathbf{Q}_x^{(\delta_x)}(0,a)\boldsymbol{x}(a)\right)\left(\sum_b \mathbf{Q}_y^{(\delta_y)}(0,b)\boldsymbol{y}(b)\right)\left(\sum_c \mathbf{Q}_z^{(\delta_z)}(0,c)\boldsymbol{z}(c)\right) \\ \left(\sum_a \mathbf{Q}_x^{(\delta_x)}(1,a)\boldsymbol{x}(a)\right)\left(\sum_b \mathbf{Q}_y^{(\delta_y)}(0,b)\boldsymbol{y}(b)\right)\left(\sum_c \mathbf{Q}_z^{(\delta_z)}(0,c)\boldsymbol{z}(c)\right) \\ \vdots \\ \left(\sum_a \mathbf{Q}_x^{(\delta_x)}(3,a)\boldsymbol{x}(a)\right)\left(\sum_b \mathbf{Q}_y^{(\delta_y)}(3,b)\boldsymbol{y}(b)\right)\left(\sum_c \mathbf{Q}_z^{(\delta_z)}(3,c)\boldsymbol{z}(c)\right) \end{bmatrix}
$$

$$(9)$$

leading to the expression

$$\mathbf{\Gamma}^{(\delta_x,\delta_y,\delta_z)} = \boldsymbol{\gamma}^{(\delta_x,\delta_y,\delta_z)} \otimes \boldsymbol{\gamma}^{(\delta_x,\delta_y,\delta_z)} \tag{10}$$

which allows for the production of the polynomial expressions for the squared second order partial derivatives by setting $(\delta_x, \delta_y, \delta_z)$ and operating directly on the control-point coefficients. For example,

$$\left(\frac{\partial^2 \nu_x}{\partial x \partial z}\right)^2 = \boldsymbol{p}_x^{\mathbf{T}}\left(\mathbf{\Gamma}^{(1,0,1)}\right)\boldsymbol{p}_x . \tag{11}$$

We can now devise a single matrix operator for computing (8) over any given region supported by a set of 64 control-points. Fig. 1 provides a 2D visualization. To later simplify computation, we separate the term $\mathbf{\Gamma}$ by B-spline basis orientation such that:

$$\mathbf{\Gamma}^{(\delta_x,\delta_y,\delta_z)} = \mathbf{\Gamma}_x^{(\delta_x)} \otimes \mathbf{\Gamma}_y^{(\delta_y)} \otimes \mathbf{\Gamma}_z^{(\delta_z)} . \tag{12}$$

By separating the 4 rows of $\mathbf{Q}_x^{(\delta_x)}$ into unit vectors

$$\mathbf{Q}_x^{(\delta_x)} = \begin{bmatrix} \boldsymbol{q}_{x,0}^{\mathbf{T}} \\ \boldsymbol{q}_{x,1}^{\mathbf{T}} \\ \boldsymbol{q}_{x,2}^{\mathbf{T}} \\ \boldsymbol{q}_{x,3}^{\mathbf{T}} \end{bmatrix}^{(\delta_x)} \tag{13}$$

we may define the sixteen 4×4 matrices given by $\mathbf{\Xi}_{x,a,b} = \boldsymbol{q}_{x,a} \otimes \boldsymbol{q}_{x,b}$ and construct the 4×4 matrix:

$$\mathbf{\Gamma}_x^{(\delta_x)}(a,b) = \mathbf{\Xi}_{x,a,b} . \tag{14}$$

Grouping like order polynomials terms within $\mathbf{\Xi}_{x,a,b}$ yields the column vector $\boldsymbol{\sigma}_{x,a,b}$:

$$\boldsymbol{\sigma}_{x,a,b} = \begin{bmatrix} \mathbf{\Xi}(0,0) \\ \mathbf{\Xi}(0,1) + \mathbf{\Xi}(1,0) \\ \mathbf{\Xi}(0,2) + \mathbf{\Xi}(1,1) + \mathbf{\Xi}(2,0) \\ \mathbf{\Xi}(0,3) + \mathbf{\Xi}(1,2) + \mathbf{\Xi}(2,1) + \mathbf{\Xi}(3,0) \\ \mathbf{\Xi}(1,3) + \mathbf{\Xi}(2,2) + \mathbf{\Xi}(3,1) \\ \mathbf{\Xi}(2,3) + \mathbf{\Xi}(3,2) \\ \mathbf{\Xi}(3,3) \end{bmatrix}_{x,a,b} \tag{15}$$

and by integrating the resulting 8th order Cartesian bases over $\boldsymbol{r}$

$$\boldsymbol{\psi}_x = \begin{bmatrix} r_x & \frac{1}{2}r_x^2 & \frac{1}{3}r_x^3 & \frac{1}{4}r_x^4 & \frac{1}{5}r_x^5 & \frac{1}{6}r_x^6 & \frac{1}{7}r_x^7 \end{bmatrix}^{\mathbf{T}} \tag{16}$$

the integral of $\mathbf{\Gamma}_x^{(\delta_x)}$ over a B-spline region may be expressed as a 4×4 matrix of vector products

$$\bar{\mathbf{\Gamma}}_x^{(\delta_x)}(a,b) = \int_0^{r_x} \mathbf{\Gamma}_x^{(\delta_x)}(a,b)dx = \boldsymbol{\sigma}_{x,a,b}^{\mathbf{T}}\boldsymbol{\psi}_x \tag{17}$$

and similarly for $\mathbf{\Gamma}_y$ and $\mathbf{\Gamma}_z$. This allows for the construction of the six desired composite matrix operators

$$\mathbf{V}^{(\delta_x,\delta_y,\delta_z)} = \begin{cases} \bar{\mathbf{\Gamma}}_x^{(\delta_x)} \otimes \bar{\mathbf{\Gamma}}_y^{(\delta_y)} \otimes \bar{\mathbf{\Gamma}}_z^{(\delta_z)} & \textbf{for} \quad \delta_x + \delta_y + \delta_z = 2 \\ 0 & \textbf{otherwise} \end{cases} \quad (18)$$

which facilitate the rapid computation of the smoothness metric over a region indexed by (l,m,n) as

$$S_{l,m,n} = \sum_{(\delta_x,\delta_y,\delta_z)} \left(\boldsymbol{p}_x^{\mathbf{T}} \mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_x + \boldsymbol{p}_y^{\mathbf{T}} \mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_y + \boldsymbol{p}_z^{\mathbf{T}} \mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_z \right) \quad (19)$$

with derivative w.r.t to a B-spline control-point $\boldsymbol{P}_{i,j,k}$

$$\frac{\partial S_{l,m,n}}{\partial \boldsymbol{P}_{i,j,k}} = \sum_{(\delta_x,\delta_y,\delta_z)} \left(2\mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_x + 2\mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_y + 2\mathbf{V}^{(\delta_x,\delta_y,\delta_z)} \boldsymbol{p}_z \right) \cdot \quad (20)$$

The total penalty S and its gradient are expressable via the summations

$$S = \sum_{(l,m,n)} S_{l,m,n} \quad \text{and} \quad \frac{\partial S}{\partial \boldsymbol{P}_{i,j,k}} = \sum_{l=0}^{3} \sum_{m=0}^{3} \sum_{n=0}^{3} \frac{\partial S_{l,m,n}}{\partial \boldsymbol{P}_{i,j,k}} , \quad (21)$$

where the summation for S indexed by (l,m,n) is over all regions and the summation for the gradient is over the 64 regions within the local support of the control point $\boldsymbol{P}_{i,j,k}$.

3   Results

We assess the performance of our analytic method by comparison to a numerical method that computes the squared second derivatives from (8) via direct central differencing of the deformation field $\boldsymbol{\nu}$, which is accumulated over the overlap domain $\boldsymbol{\Omega}$. Computational time required by such an approach is proportional to the number of voxels within $\boldsymbol{\Omega}$. By contrast, the time required by the proposed analytic method is proportional to the number of regions defined by the B-spline control-point spacing; thus reducing the complexity. Furthermore, the

Table 1. Wall clock execution times and associated speed-ups for the proposed analytic scheme vs numerical central differencing of the vector field

| | | Processing Time | | |
Volume Size	Control-Point Spacing	Numeric	Analytic	Speed-up
$256 \times 256 \times 256$	$10 \times 10 \times 10$	91.746s	1.505s	61x
$256 \times 256 \times 256$	$30 \times 30 \times 30$	90.238s	0.126s	722x
$512 \times 512 \times 512$	$10 \times 10 \times 10$	758.902s	11.650s	65x
$512 \times 512 \times 512$	$30 \times 30 \times 30$	762.041s	0.556s	1371x

(a) Image difference   (b) $\mathbf{T}$ for $\lambda = 0$ (c) $\mathbf{T}$ for $\lambda = 0.007$ (d) $\mathbf{T}$ for $\lambda = 0.7$
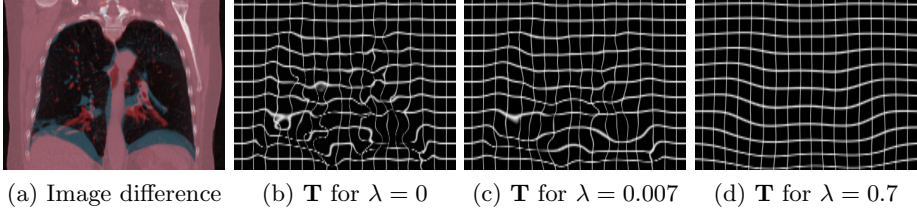
Fig. 2. Coronal slice of of deformation field $\mathbf{T}$ acquired by registering (a) exhaled and inhaled thoracic CT images. $\mathbf{T}$ is computed with a $10 \times 10 \times 10$ control-point spacing for (b-d) different values of λ.

six $\mathbf{V}$ matrices may be pre-computed and reused throughout the optimization of the B-spline coefficients $\mathbf{P}$ for a given control-point spacing $\boldsymbol{r}$. As shown in Table 1, our analytic method achieves a speed-up ranging between 61x–1371x, dependent upon the voxel to region ratio for a given registration configuration. Agreement between numerical and analytic solutions is within 2% [14]. The effect of the bending-energy penalty factor λ on the transform $\mathbf{T}$ is qualitatively demonstrated by Fig. 2 where inhaled and exhaled thoracic volumes with displacement shown by Fig. 2(a) are registered using a B-spline control-point spacing of $10 \times 10 \times 10$ mm for λ varying over several orders of magnitude. As expected, the resulting transforms $\mathbf{T}$ shown in Fig. 2(b-d) exhibit decreasing bending energy for increasing λ. This single example is from one of ten case studies performed to quantitatively explore the effectiveness of our analytic regularization as a function of control-point spacing $\boldsymbol{r}$ and penalty factor λ.

Each of the ten studies consists of an image volume taken at full inhalation and a subsequent volume at full exhalation. Five of the ten studies have volumes of $512 \times 512 \times 128$ voxels with physical separations of $0.92 \times 0.92 \times 2.5$ mm; the remaining five studies have the same physical separations but volumes are lower resolution at $256 \times 256 \times 128$ voxels. For each image, 300 landmarks are placed within the lung by a medical expert. Registrations are performed using the mean-squared error similarity metric penalized by S with weight λ as in (1). The B-spline coefficients $\mathbf{P}$ describing the transform $\mathbf{T}$ are optimized via the quasi-Newtonian method implemented by the L-BFGS-B optimizer using an analytically computed cost function and gradient. Fig. 3(a) shows the mean separation of corresponding landmarks between inhaled and exhaled volumes as a function of control-point spacing and λ after application of the computed transform $\mathbf{T}$. The primary range of interest falls mostly within 10^{-1}–10^{2} where increasing values of λ produce increasing mean separations between landmarks. As shown in Fig. 3(c), the minimum Jacobian determinate of $\mathbf{T}$ reveals that as λ increases through this range, the resulting transforms tend to increase in smoothness until becoming nearly plastic as $|\nabla \mathbf{T}|$ approaches unity within the range $1 < \lambda < 2$. In the absence of regularization, control-point spacings 10–30mm tend to produce non-smooth local deformations due to the restricted influence of the B-spline basis; however, the resulting increased parameterization

(a) Mean landmark separation

(b) min $|\nabla \boldsymbol{T}|$
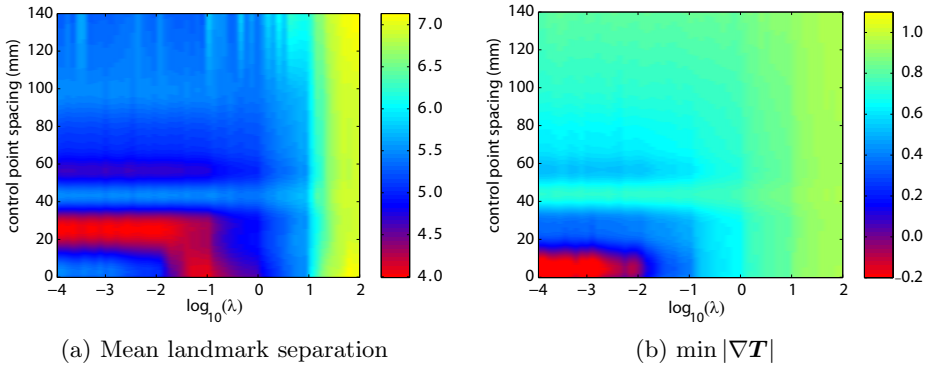
Fig. 3. Average separation of corresponding landmarks as a function of B-spline control-point spacing and the bending energy penalty factor λ over the 10 thoracic cases.

of $\mathbf{T}$ allows for more complex local variations to be described. Small values of λ less than 10^{-2} have little effect in this region and $|\nabla \mathbf{T}|$ deviates greatly from unity. This is an indication of aggressive local expansion and compression in $\mathbf{T}$ that is most likely physiologically unsound. Furthermore, the occurrence of negative values for $|\nabla \mathbf{T}|$ indicate the presence of non-orientation preserving local transforms within this range of λ. Consequently, increasing λ beyond 10^{-2} exhibits the unique quality of increasing registration accuracy as well as the potential physiological applicability of $\mathbf{T}$ for such fine control-point spacings, as shown by Fig. 3(a).

4 Conclusions

We have developed an analytically derived set of composite matrix operators that operate directly on a set of 64 control-points to produce the bending energy within a given region of support. The behavior of our method has been characterized by application to ten thoracic studies and it was demonstrated that our method of bending energy computation provides a speed-up within the range of 60–$1371x$ depending on input volume resolution and B-spline control-point spacing.

This algorithm has been implemented as part of Plastimatch and can be downloaded under a BSD-style license from http://www.plastimatch.org.

References

1. Scahill, R., Frost, C., Jenkins, R., Whitwell, J., Rossor, M., Fox, N.: A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. Arch. Neurol. 60, 989–994 (2003)
2. Job, D., Whalley, H., McConnell, S., Glabus, M., Johnstone, E., Lawrie, S.: Voxel-based morphometry of grey matter densities in subjects at high risk of schizophrenia. Schizophr Res. 64, 1–13 (2003)
3. Thompson, P., Giedd, J., Woods, R., MacDonald, D., Evans, A., Toga, A.: Growth patterns in the developing human brain detected using continuum-mechanical tensor mapping. Nature 404, 190–193 (2000)
4. Boctor, E., deOliveira, M., Choti, M., Ghanem, R., Taylor, R., Hager, G., Fichtinger, G.: Ultrasound Monitoring of Tissue Ablation Via Deformation Model and Shape Priors. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 405–412. Springer, Heidelberg (2006)
5. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
6. Flampouri, S., Jiang, S., Sharp, G., Wolfgang, J., Patel, A., Choi, N.: Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. Phys. Med. Biol. 51(11), 2763–2779 (2006)
7. Zhang, T., Chi, Y., Meldolesi, E., Yan, D.: Automatic delineation of on-line head-and-neck computed tomography images: Toward on-line adaptive radiotherapy. Int. J. Radiat. Oncol. 68(2), 522–530 (2007)
8. Rohkohl, C., Lauritsch, G., Prümmer, M., Hornegger, J.: Interventional 4-D motion estimation and reconstruction of cardiac vasculature without motion periodicity assumption. Med. Image Anal. (2010)
9. Brunet, T., Nowak, K., Gleicher, M.: Integrating dynamic deformations into interactive volume visualization. In: Eurographics/IEEE VGTC Symposium on Visualization 2006, Citeseer, pp. 219–226 (2006)
10. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. IEEE T. Med. Imaging 18(8), 712–721 (1999)
11. Rohlfing, T., Maurer Jr., C., Bluemke, D., Jacobs, M.: Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. IEEE T. Med. Imaging 22(6), 730–741 (2003)
12. Miller, M., Christensen, G., Amit, Y., Grenander, U.: Mathematical textbook of deformable neuroanatomies. P. Natl. Acad. Sci. USA 90(24), 11944 (1993)
13. Li, X., Dawant, B., Welch, E., Chakravarthy, A., Freehardt, D., Mayer, I., Kelley, M., Meszoely, I., Gore, J., Yankeelov, T.: A nonrigid registration algorithm for longitudinal breast MR images and the analysis of breast tumor response. Magn. Reson. Imaging 27(9), 1258–1270 (2009)
14. Yang, Q.: Analytic Regularization for B-spline Deformable Image Registration. Master's thesis, der Universitat Heidelberg, Germany (2011)

Simultaneous Multiscale Polyaffine Registration by Incorporating Deformation Statistics

Christof Seiler[1,2], Xavier Pennec[2], and Mauricio Reyes[1]

[1] Institute for Surgical Technology and Biomechanics,
University of Bern, Switzerland
[2] Asclepios Research Group, INRIA Sophia Antipolis, France

Abstract. Locally affine (polyaffine) image registration methods capture intersubject non-linear deformations with a low number of parameters, while providing an intuitive interpretation for clinicians. Considering the mandible bone, anatomical shape differences can be found at different scales, e.g. left or right side, teeth, etc. Classically, sequential coarse to fine registration are used to handle multiscale deformations, instead we propose a simultaneous optimization of all scales. To avoid local minima we incorporate a prior on the polyaffine transformations. This kind of groupwise registration approach is natural in a polyaffine context, if we assume one configuration of regions that describes an entire group of images, with varying transformations for each region. In this paper, we reformulate polyaffine deformations in a generative statistical model, which enables us to incorporate deformation statistics as a prior in a Bayesian setting. We find optimal transformations by optimizing the maximum a posteriori probability. We assume that the polyaffine transformations follow a normal distribution with mean and concentration matrix. Parameters of the prior are estimated from an initial coarse to fine registration. Knowing the region structure, we develop a blockwise pseudoinverse to obtain the concentration matrix. To our knowledge, we are the first to introduce simultaneous multiscale optimization through groupwise polyaffine registration. We show results on 42 mandible CT images.

1 Introduction

Mandibular fractures most commonly result from facial trauma, with close to half of the patients requiring surgical repair [6]. A majority of 75% of fractures occur in males aged between 20 and 30 [6], and are often caused by physical assault. For these cases surgical repair proofs most effective, with the goal of recovering the anatomical structure prior to the injury and thus restoring normal function. To reach this goal the surgeon places wires or implants at the fracture site, so that the natural fusion of separated bone pieces restores the prior structure as closely as possible. A correct repair aligns teeth for food intake, and restores the patient's aesthetics. In [12], the authors propose a classification scheme for mandibles based on regions according to anatomical, functional and aesthetic considerations. The online register www.aofoundation.org uses the
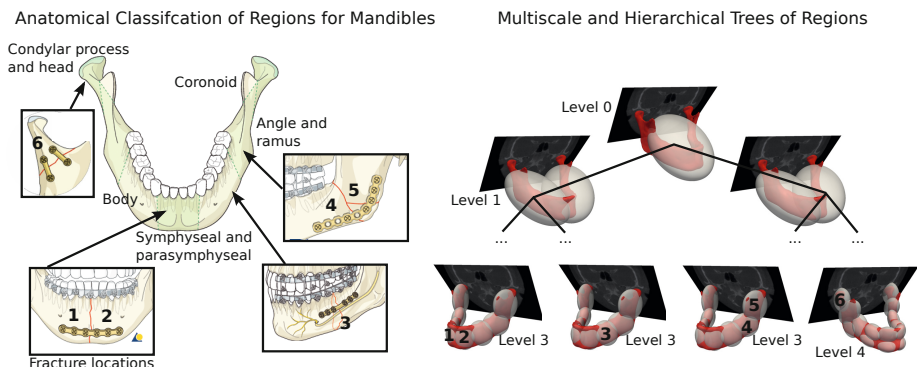
Fig. 1. Left: Subdivision of mandible into anatomical regions proposed by the AO foundation to classify fractures for reconstructive surgery. Implants at four different anatomical sites are shown. Images source: www.aofoundation.org. **Right:** Tree of Gaussian weights. Each Gaussian weight represents one region and is visualized as one ellipsoids at σ. The contours extracted from CT data in red and one slice of the original CT image (to indicate that we work in the image domain and not only on the contour) are shown.

same classification (Fig. 1) to guide surgeons through the major steps of mandible reconstructive surgery ranging from diagnosis, selection of the optimal surgical approach, to aftercare treatment. In addition to this classification scheme, the mandible can be subdivided even further into smaller regions, e.g. one region for each tooth. In the image space, this subdivision can theoretically be performed up to the voxel level, where coarser levels enclose finer ones, representing a hierarchy of regions that can be organized in a tree-like fashion. As the regions become more fine, it is harder to find a consensus among clinicians on the size, shape and location of the region.

Recent work on biomechanical analysis of implants indicates that geometry and topology of implants are crucial to fracture stability [7]. Due to the high economical cost of patient specific implants current approaches focus on population-based implant design. Common key steps to population-based design, e.g. [3], are registration to capture shape variability as encountered in a population and statistical analysis of the registration results, performed subsequently and independently.

Due to region-based description of the mandible shape and the need for volumetric information we focus on locally affine transformations, also called polyaffine transformations. Polyaffine transformations fuse locally rigid and affine transformations into a diffeomorphism [1]. An efficient registration algorithm using approximations of polyaffine transformations was presented in [5]. To consider more complex shapes and foster reusability, [4] presented a multiscale approach with affine regions defined using a data-driven approach. The method splits rectangular shaped regions, which are aligned along the image directions, only if certain conditions are met. The authors in [11] iteratively optimize between affine parameters and anchor positions (center of regions) estimation,

through an expectation maximization approach. Even though these methods are very promising, the link between the clinical regions (Fig. 1) and the regions found by these algorithms, is either constrained by aligned rectangular shaped regions [4] or produce an intractable number (around 500) of regions [11]. In [10], the authors introduced a hierarchical multiscale tree structure (called polyaffine transformation trees, or PolyTree in short) that is motivated by the nature of the mandible anatomy, where regions are ordered and interact with each other in a way tractable for human understanding. Aforementioned polyaffine methods are pairwise registrations, and to our knowledge, no groupwise registration (e.g. [9,8]) has been proposed in this context. We believe that transformations on a groupwise level are crucial. This is motivated by the assumption that there should be one configuration of regions that describes an entire group of images, with varying transformations for each region. Furthermore, the multiscale property of mandibles and other anatomical structures should be incorporated.

In [2], the authors presented a Bayesian approach for affine registration, in this paper, we propose a Bayesian formulation of polyaffine registration across scales. We reformulate PolyTrees in a probabilistic way. We introduce a generative statistical model, which enables us to incorporate deformation statistics as a prior in a Bayesian setting. We find optimal transformations by optimizing the maximum a posteriori probability (MAP). In Section 2, we describe PolyTrees, which were recently introduced in [10], to describe intersubject deformations. In Section 3, we reformulate PolyTrees in a probabilistic way and show how to find optimal transformations using MAP estimates with groupwise deformation statistics as a prior. In Section 4, we show results on 42 mandible CT images.

2   Multiscale Description of Intersubject Deformations

In this section we reinterpret polyaffine transformation trees (PolyTree) recently introduced in [10]. PolyTrees are parametric transformation that describe non-rigid deformations with a low number of parameters and captures the shape variability of multiscale structures. First, we introduce the tree structure, second we define data-driven regions, and lastly, we show the estimation of transformations using the log-demons. The difference with [10] is that we reformulate the equations in vectorized form: Vect is the column-wise vectorization of a matrix $m = \text{Vect}(M)$, where m is a vector and M a matrix. The motivation for this will become evident in Section 3.

Structuring of Locally Affine Transformations in Trees. For N regions let $M = \begin{bmatrix} M_1 \ldots M_N \end{bmatrix}^T$ be the $3 \times 4N$ non null components of the matrix logarithm of affine transformations and $v_M(x)$ be the parametrized polyaffine stationary velocity field:

$$v_M(x) = M(w(x) \otimes \tilde{x}), \;\; \log\left(\begin{bmatrix} A_i & t_i \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} M_i \\ 0 \end{bmatrix}, \tag{1}$$

where $\tilde{x} = \begin{bmatrix} x & 1 \end{bmatrix}^T$, x is the spatial position, $w(x) = \begin{bmatrix} w_1(x) \ldots w_N(x) \end{bmatrix}^T$, $\otimes$ is the Kronecker product, A_i is the affine transformation matrix and t_i is the

translation vector. The weights are normalized, $\forall x \in \Omega : \sum_{i=1}^{n} w_i(x) = 1$, and structured in a tree-like fashion as depicted in Fig. 1, this enables to describe multiscale deformations.

Data-Driven Definition of Regions. The spatial weight functions $w_i(x)$ are defined as multivariate Gaussian distributions, $w_i(x) = \mathcal{N}(\bar{x}_i, \Xi_i)$. The parameters of the distributions are found by applying the oriented bounding box algorithm (OBBTree) to the feature image, $\psi(x) = \log(1 + ||\nabla I_t||^2)$, where ∇I_t is the gradient of template image. We take the logarithm of the gradient to be more robust against small changes in intensities due to noise. The feature-weighted barycenter of region Ω_i (discretized at voxel indices j), $\bar{x}_i = (\sum_{j \in \Omega_i} \psi(x_j))^{-1} \sum_{j \in \Omega_i^l} \psi(x_j) x_j$, and the feature-weighted covariance matrix of the region, $\Xi_i = (\sum_{j \in \Omega_i^l} \psi(x_j))^{-1} \sum_{j \in \Omega_i^l} \psi(x_j)(x_j - \bar{x}_i^l)(x_j - \bar{x}_i^l)^T$. Following the same strategy as in the original OBBTree algorithm, the region is split at the feature-weighted mean point orthogonal to the first principle component and the splitting procedure is recursively repeated in the two new created subregions. This step is identical to the method proposed in [10].

Estimation of Transformations with the Log-Demons Algorithm. The general form of the log-demons algorithm for stationary velocity fields (SVF) was formulated in [13]. The goal is to find v_M that warps the fixed image I_t into the moving image I_s (or resamples I_s in I_t), by minimizing the cost functional, $C(I_t, I_s, v, v_M) = \sigma_i^{-2} \text{Sim}(I_t, I_s, v_M) + \sigma_x^{-2} \text{dist}(v, v_M)^2 + \sigma_T^{-2} \text{Reg}(v)$, where Sim, dist and Reg, are the similarity, the hidden and the regularization term, respectively. Each term has a weighting parameter σ_i, σ_x and σ_T. As shown in [13], $C(I_t, I_s, v, v_M)$ can be optimized alternatively over the variables v and v_M. Given v, the correspondence velocity field computed by the first optimization part of the log-demons algorithm, we solve for M using linear least squares, i.e. minimizing $C(M) = \int_\Omega \lambda(x) ||v(x) - M(w(x) \otimes \tilde{x})||^2 dx$, where λ is a binary mask indicating background voxels (if no mask is available $\forall x \in \Omega : \lambda(x) = 1$). In contrast to the original formulation in [10] the transformations parameters M are vectorized, and the Kronecker product is used. This reformulation is necessary for the next section, where we present a generative statistical model, which enables us to incorporate deformation statistics as a prior in a Bayesian setting. We find optimal transformations by optimizing the MAP.

3 Incorporating Deformation Statistics

In this section we present the MAP to find optimal transformations with and without prior for K patients. We denote $\hat{m}_k$ as the kth transformation found with registration without prior, and $\check{m}_k$ with prior. We denote $[k]$ to consider all patients.

Generative Statistical Model of Polyaffine Transformation Trees. Here we make two assumptions: First, we assume that the velocities are independent

at all voxel positions within the mask Ω. This is an approximation in the case of the log-demons algorithm, where the voxels of correspondence field v are not entirely independent due to the natural smoothness of the images used. Second, we assume the independence of the velocities across subjects. This is only the case if the template with which the patients are registered is unbiased. With these assumption we end up with the log likelihood (modulo unnecessary constants),

$$\mathcal{L}(v^{[k]}|m^{[k]}) = -\sum_k \log(P(v^k|m^k)) = \frac{1}{2\sigma_v^2} \sum_k^K \int_\Omega \lambda(x)||v^k(x) - v_{M^k}(x)||^2 dx, \tag{2}$$

with the underlying Gaussian noise model $v^k(x) = v_{M^k}(x) + \mathcal{N}(0, \sigma_v^2 I_3)$, where I_3 is the 3×3 identity matrix, and K subjects. For one patient, the probability to observe v^k given m^k is,

$$P(v^k|m^k) \propto \exp\left(-\frac{1}{2}(m^k - \hat{m}^k)^T \Sigma (m^k - \hat{m}^k)\right), \tag{3}$$

where $\hat{m}^k$ denotes the optimal transformation parameters without a prior on m^k. To find the optimal solution we take the derivative of $\mathcal{L}(v^k|m^k)$ with respect to m^k and set it to zero,

$$\hat{m}^k = \Sigma^{-1} b^k, \tag{4}$$
$$\Sigma = \frac{1}{2\sigma_{v^2}} \int_\Omega \lambda(x) \left(w(x)w(x)^T\right) \otimes (\tilde{x}\tilde{x}^T) dx \otimes I_3,$$
$$b_k = \frac{1}{2\sigma_{v^2}} \int_\Omega \lambda(x) \left(w(x) \otimes \tilde{x}\right) \otimes v^k(x) dx,$$

where dx is a scalar.

Maximum Likelihood Estimation with Prior on the Transformations.
Now we show how to find $\check{m}^k$ transformations with a prior. Assume that $m^k \sim \mathcal{N}(\bar{m}, \Gamma)$, where $\bar{m}$ is the mean and Γ the concentration matrix. The probability of observing a transformation given the prior distribution of m^k is,

$$P(m^k|\theta) = \frac{\det(\Gamma)^{1/2}}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(m^k - \bar{m})^T \Gamma (m^k - \bar{m})\right), \tag{5}$$

where $d = 12N$. Assume that we know $\theta = (\bar{m}, \Gamma)$ and that we want to register with this prior, for this we optimize the MAP estimate,

$$P(m^k|v^k, \theta) = \frac{P(v^k|m^k, \theta)P(m^k|\theta)}{P(v^k|\theta)}, \tag{6}$$

the optimal solution of $\mathcal{L}(m^k|\theta)$ (modulo the constant $P(v^k|\theta)$) is,

$$\check{m}^k = (\Sigma + \Gamma)^{-1}(\Sigma\hat{m}^k + \Gamma\bar{m}). \tag{7}$$

For our experiments on mandible CT images, we estimate Γ using a blockwise pseudoinverse described in the next paragraph.

Blockwise Pseudoinverse of Sample Covariance Matrix for Prior. To estimate Γ we take advantage of the region structure, which reflects as a blockwise structure in the covariance matrix, $\text{Cov} = \frac{1}{K}\sum_k^K (\hat{m}^k - \bar{m})(\hat{m}^k - \bar{m})^T$, where $\bar{m}$ is the sample mean. We develop a blockwise pseudoinverse, $\Gamma = \text{diag}(\lambda_i^+ \otimes C^{-1})$, where $.^+$ is the pseudoinverse, λ_i is the variance of block i (i.e. region i), and C is a 12×12 matrix, representing the metric of the log affine parameters. We optimize, $\{\hat{C}, \hat{\lambda}_i\} = \text{argmin}_{C,\lambda_i} = \sum_i^N \| \text{Cov}_{ii} - \lambda_i C\|^2$, in two sequentially steps. In step one, for C with fixed $\forall i : \lambda_i = 1$, and in step two for λ_i with fixed C obtained from the previous step:

$$\text{Step one: } \hat{C} = \frac{\sum_{i=1}^N \lambda_i \text{Cov}_{ii}}{\sum_{i=1}^N \lambda_i^2}, \qquad \text{Step two: } \hat{\lambda}_i = \frac{\text{Trace}(\text{Cov}_{ii}\, C)}{\text{Trace}(C^2)}. \qquad (8)$$

4   Experiments on Mandible CT Image Data

In this section, we register 42 mandible CT images to a randomly chosen template and evaluate the resulting transformations. The noise parameter is set to $\sigma_v = 1.85$ mm, and the number of affine components is $N = 31$ across 5 levels. To evaluate the influence of the prior, we run five different experiments: no prior, λ_i scaled by 1,0.1,0.01 and 0.001. For the evaluation of the robustness we decompose the log affine transformations into three parts: rotation, expansion and translation. The decomposition of the affine part into skew symmetric
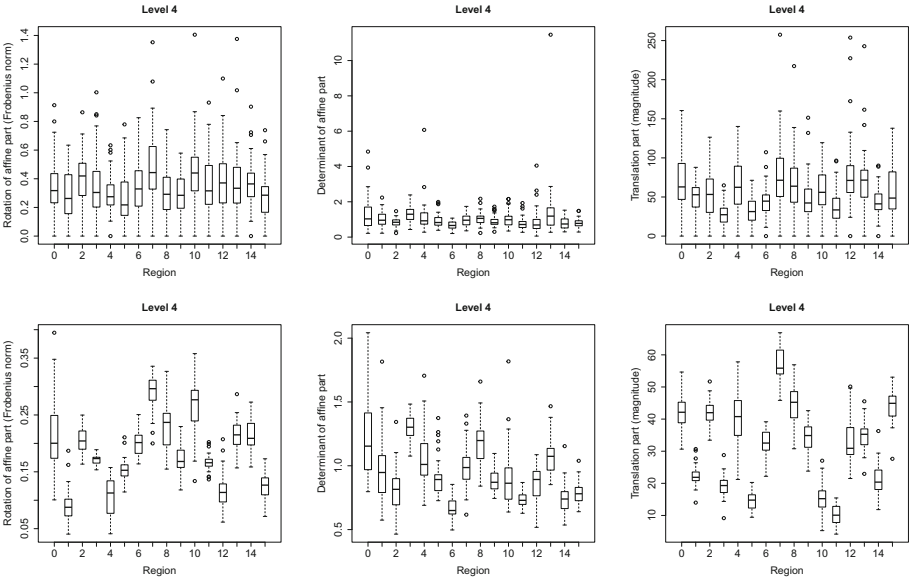
Fig. 2. In **top row** statistics on transformations obtained with registration without prior and **bottom row** with prior for level 4 (i.e. 16 regions)

and symmetric matrices (representing rotation and expansion, respectively) is, $M_i = \frac{1}{2}(M_i - M_i^T) + \frac{1}{2}(M_i + M_i^T)$, and the Jordan or Schur decomposition to compute the determinant of an affine part, is given by, $\det(A_i) = \exp(\text{Trace}(M_i))$.
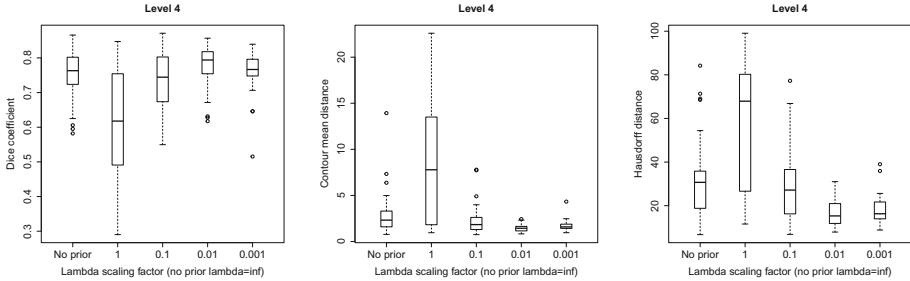
Fig. 3. Comparison of registration accuracy in terms of Dice coefficient, contour mean distance and Hausdorff distance (both in mm) at level 4. The parameter λ_i was scaled with five different factors to evaluate the influence of the prior.

The robustness results are shown for level 4 (i.e. 16 regions) in Fig. 2. The registrations without prior show outliers and very high expansion of regions up to a determinant of 10. In the registration with prior (the prior was scaled by 0.01), we observe robust expansion and shrinkage factors between 0.5 and 2. The accuracy of the registration is measured in terms of Dice coefficient of the mask images (semi-manual segmented mask are available for all 42 images), mean contour distance, and Hausdorff distance on the mask contour. In Fig. 3, by scaling the prior with 0.01 all accuracy measure show a favorable median value compared to the without prior registration.

5   Conclusion

In this work, we presented a multiscale polyaffine registration method that simultaneously registers all scales. To avoid local solutions we incorporated deformations statistics. The results showed that the registration with prior improves the accuracy while reducing the variability of estimated transformation parameters.

We believe that our approach is more than just an extension of [2]. By considering not only one affine component, but a mixture of components acting at different scales, we are moving the discussion into structured learning, which to our knowledge is a novelty in the medical registration community. In future work, we plan to introduce a sparse representation of anatomical substructures and their connection at different scales, which might uncover structures equivalent to rigid articulated bodies.

References

1. Arsigny, V., Commowick, O., Ayache, N., Pennec, X.: A Fast and Log-Euclidean Polyaffine Framework for Locally Linear Registration. J. Math. Imaging Vis. 33(2), 222–238 (2009)
2. Ashburner, J., Neelin, P., Collins, D.L., Evans, A., Friston, K.: Incorporating Prior Knowledge Into Image Registration. NeuroImage 6(4), 344–352 (1997)
3. Bou-Sleiman, H., Ritacco, L.E., Nolte, L.-P., Reyes, M.: Minimization of Intra-Operative Shaping of Orthopaedic Fixation Plates: A Population-Based Design. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 409–416. Springer, Heidelberg (2011)
4. Buerger, C., Schaeffter, T., King, A.P.: Hierarchical Adaptive Local Affine Registration for Fast and Robust Respiratory Motion Estimation. Med. Image Anal. 15(4), 551–564 (2011)
5. Commowick, O., Arsigny, V., Isambert, A., Costa, J., Dhermain, F., Bidault, F., Bondiau, P.Y., Ayache, N., Malandain, G.: An Efficient Locally Affine Framework for the Smooth Registration of Anatomical Structures. Med. Image Anal. 12(4), 427–441 (2008)
6. Ellis, E., Moos, K.F., El-Attar, A.: Ten Years of Mandibular Fractures: An Analysis of 2,137 Cases. Oral Surg. Oral Med. Oral Pathol. 59(2), 120–129 (1985)
7. Lovald, S., Baack, B., Gaball, C., Olson, G., Hoard, A.: Biomechanical Optimization of Bone Plates Used in Rigid Fixation of Mandibular Symphysis Fractures. Journal of Oral and Maxillofacial Surgery 68(8), 1833–1841 (2010)
8. Marsland, S., Twining, C.J., Taylor, C.J.: A Minimum Description Length Objective Function for Groupwise Non-Rigid Image Registration. Image Vision Comput 26(3), 333–346 (2008)
9. Pennec, X., Stefanescu, R., Arsigny, V., Fillard, P., Ayache, N.: Riemannian Elasticity: A Statistical Regularization Framework for Non-linear Registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 943–950. Springer, Heidelberg (2005)
10. Seiler, C., Pennec, X., Reyes, M.: Capturing the Multiscale Anatomical Shape Variability with Polyaffine Transformation Trees. Med. Image Anal. (2012), http://dx.doi.org/10.1016/j.media.2012.05.011
11. Taquet, M., Macq, B., Warfield, S.K.: Spatially Adaptive Log-Euclidean Polyaffine Registration Based on Sparse Matches. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 590–597. Springer, Heidelberg (2011)
12. Urken, M.L., Weinberg, H., Vickery, C., Buchbinder, D., Lawson, W., Biller, H.F.: Oromandibular Reconstruction Using Microvascular Composite Free Flaps. Report of 71 Cases and a New Classification Scheme for Bony, Soft-Tissue, and Beurologic Defects. Arch. Otolaryngol Head Neck Surg. 117(7), 733–744 (1991)
13. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic Demons: Efficient Non-Parametric Image Registration. NeuroImage 45(1, suppl. 1), S61–S72 (2009)

Fast Diffusion Tensor Registration with Exact Reorientation and Regularization[*]

Junning Li[1], Yonggang Shi[1], Giang Tran[3], Ivo Dinov[1],
Danny J.J. Wang[2], and Arthur W. Toga[1]

[1] Laboratory of Neuro Imaging,
[2] Brain Mapping Center, Department of Neurology
[3] Department of Mathematics
University of California, Los Angeles, CA, USA
{Junning.Li,yshi,ivo.dinov,JJ.Wang,toga}@loni.ucla.edu
giangtran@math.ucla.edu

Abstract. Diffusion tensor imaging is widely used in brain connectivity study. As more and more group studies recruit a large number of subjects, it is important to design registration methods that are not only theoretically rigorous, but also computationally efficient, for processing large data sets. However, the requirement of reorienting diffusion tensors complicates and slows down the registration, especially for those methods whose scalar-image versions have linear complexity, for example, the Demons algorithm. In this paper, we propose an extension of the Demons algorithm that incorporates exact reorientation and regularization into the calculation of deforming velocity, yet preserving its linear complexity. This method restores the computational efficiency of the Demons algorithm to diffusion images, but does not sacrifice registration goodness. In our experiments, the new algorithm achieved state-of-art performance at a ten-fold decrease of computational time.

1 Introduction

Water molecules in biology tissue tend to diffuse faster along, relative to across, obstacle structures. Diffusion tensor imaging (DTI) noninvasively measures this anisotropy (approximated with second-order tensors), providing information about structures such as neural fibers. As more and more large group studies, such as those sponsored by the Human Connnectome Project [1], require nonlinear normalization of diffusion tensor images, it is important to design deformation registration methods that are not only theoretically rigorous, but also computationally efficient, for processing large data sets.

Registration of diffusion tensor images is more complicated than scalar images, because displacing a voxel to a new place does not only change its own diffusion tensor value, but also reorients those of its adjacent voxels [2]. As a result, the deformation force is no longer independent for adjacent voxels, but has a sparse 3D grid structure.

Because reorientation significantly complicates the computation, methods based on rotation-invariant features have been developed. Many of these methods first extract scalar-valued or vector-valued features from tensors, for instance, the fractional anisotropy (FA), and then register diffusion images by aligning these features with multi-channel registration [3,4].

Some other methods directly work on tensors, with similarity metric defined on tensors, and indirect or direct involvement of reorientation into velocity field optimization. Alexander and Gee [5] reorientated tensors according to the new displacement field after each iteration, though not directly accounted the reorientation into optimization. Zhang *et al.* [6] applied piece-wise affine registration to image subblocks, and then fused these transformations together by smoothing. Since finite-strain (FS) reorientation [2] can be analytically incorporated into affine transformation, the method efficiently estimates the optimal local affine parameters, but it is not clear how the fusion step affects the total registration energy. Cao *et al.* [7] extended the Large Deformation Diffeomorphic Metric Mapping framework, to analytically embed the preservation-of-principal-direction (PPD) reorientation [2] into the deforming force. In 2009, Yeo *et al.* [8] embedded exact FS reorientation into the diffeomorphic Demons algorithm [9,10], showing that exact reorientation considerably improved registration accuracy.

Yeo *et al.*'s work [8] was based on the Demons algorithm, a fast algorithm with $O(n)$ complexity for scalar images, where n is the number of voxels. However, because a large sparse linear system was solved at every iteration, the algorithm became considerably slower. This effect becomes more exacerbated for such an algorithm whose scalar-image version enjoys linear complexity. For example, for our DTI images of size 128x128x128, the method took about 7 hours and more than 20 Giga Bytes (GB) memory on a desktop with an Intel Xeon 2.80 GHz CPU. On the other hand, the regularization on the displacement field was separated as a Gaussian smoothing step after updating the displacement field. This may not be always consistent with the diffeomorphic framework [10].

This raises the question "can the Demons algorithm still enjoy its $O(n)$ complexity when the deformation force is coupled between adjacent voxels and regularization is incorporated?" Though this question is raised from diffusion tensor images, it is generally applicable to other diffusion images, such as high-angular-resolution diffusion images (HARDI) and diffusion spectrum images (DSI). We are particularly interested in the Demons algorithm, because its original linear complexity in [9,10] is well suited for processing large group data sets.

In this paper, we extend the Demons algorithm to incorporate both exact reorientation and regularization into velocity calculation, but without directly solving a large non-separable linear system. This method restores the $O(n)$ computational efficiency of the original Demons algorithm to diffusion images, but does not sacrifice registration goodness, in comparison with solving a large linear system at each iteration. In our experiments, it introduced a 10-fold reduction of the computation time, and achieved state-of-art registration performance.

2   Method

2.1   Diffeomorphic Demons Registration for Diffusion Tensors

Let F and M respectively denote the fixed and moving image, u and v respectively denote the transformation field and velocity field, $\circ$ denote the compositive operator, $\bullet(i)$ denote the image value at voxel i, $\bullet_i$ and $\bullet_x$ denote the velocity, displacement or a certain value associated respectively with a voxel i or a point x. The diffeomorphic Demons algorithm for scalar images in [10] is as follows:

Repeat   until convergence

1. **For each** voxel i, at iteration k, given current transformation field u^k:
   (a) **Define** $b_i(v) = F(i) - M \circ [u^k \circ \exp(v)](i)$.
   (b) **Let** $b_i = b_i(v)\,|_{v=0} = F(i) - M \circ u^k(i)$, and $g_i = \frac{\partial b_i(v)}{\partial v_i}\,|_{v=0}$.
   (c) **Let** $v_i = \frac{g_i b_i}{g_i' g_i + 1/\delta_i^2}$, where $1/\delta_i^2$ is the velocity regularizer at voxel i.
2. **Update displacement field** with the velocity field: $u^{k+1} = u^k \circ \exp(v)$.

The algorithm minimizes the energy function

$$\mathcal{E}(u) = \int [F(x) - M \circ u(x)]^2 dx + \int \left\| \frac{v_x}{\delta_x} \right\|^2 dx \tag{1}$$

with the displacement-velocity relationship $v = \frac{du}{dt} = \frac{\partial u}{\partial t} + (\nabla u)v$ (actually discretized as $u^{k+1} = u^k \circ \exp(v)$), and a discrete time interval $\tau = 1$. Smoothing can be applied to the displacement field or the velocity field, to impose regularization. The minimization is by approximating the energy function as $\mathcal{E} = \sum_i \left\| (\sum_j \frac{\partial b_i(v)}{\partial v_j} v_j)\,|_{v=0} - b_i \right\|^2 + \sum_i \left\| \frac{v_i}{\delta_i} \right\|^2$ with the Gauss-Newton method. For scalar images, $\frac{\partial b_i(v)}{\partial v_j}\,|_{v=0}$ equals zero if $i \neq j$, so the optimization can be separated for each voxel as in Step (1) and has $O(n)$ complexity at each iteration, where n is the number of voxels. For details, please refer to [9,10].

For diffusion tensor images, tensors must be reorientated during transformation to align their directions with the transformed space. Consequently, the displacement at a voxel not only changes its own warped tensor value, but also reorientates those of its adjacent voxels, that is, $\frac{\partial b_i(v)}{\partial v_j}\,|_{v=0}$ is not zero for adjacent voxels i and j. In this case, the energy function cannot be separated for each voxel. For details on the formulas of tensor reorientation, please refer to [2].

2.2   Regularization

Standard regularizers for displacement fields include the elastic, diffusion, and curvature regularizers [11] which all can be formatted as the squared norm of a certain derivative value of the displacement field. When discretized, these regularizers can be formulated as $\sum_{i=1}^n \|A_{\mathrm{reg}\,i} u\|^2$ where $A_{\mathrm{reg}\,i}$ is a differential operator at voxel i, represented as a matrix. In this paper, we used $\sum_l \int \|H_l u\|^2 dx$ as the regularizer, where H_l is the Hessian operator applied to the lth dimensional component of u. We call it affineness regularizer because the affine transformation is the only non-trivial kernel not punished by this regularizer.

2.3   Local Gauss-Newton Optimization

As shown in Sections 2.1 and 2.2, the energy function of the Demons algorithm, including both the dissimilarity and regularization terms, is the sum of squares of many linear functions. Therefore, the energy function can be written as $\mathcal{E} = \|A_{img}v - b_{img}\|^2 + \|A_{reg}v - b_{reg}\|^2 = \|Av - b\|^2$. Due to tensor reorientation, A is not separable for adjacent voxels, and the optimization cannot be separated as in Step (1) of the Demons algorithm for scalar images. In [8], Yeo minimized the energy by solving this large sparse linear system at each iteration, excluding the regularization term. Though the system is sparse, the computational cost still goes up sharply as the image size increases. For images of size 128x128x128, we observed 20 GB memory usage and about 7 hours computation. This raises the question "can the Demons algorithm still have its $O(n)$ complexity when the deformation force is coupled between adjacent voxels?"

In this paper, we propose a method that incorporates both tensor reorientation and displacement regularization into its optimization iterations, while restores the $O(n)$ complexity of the Demons algorithm. The algorithm is as follows, where we rewrite the energy function as $\mathcal{E} = \|Av - b\|^2 = \|(\sum A_i v_i) - b\|^2$ and A_i is the columns of A related to the velocity at voxel i:

1. **For each voxel** i:
   (a) If $\|A_i' b\| <$ stop threshold, then let $d_i = 0$, else let $d_i = \underset{z}{\operatorname{argmin}}\|A_i z - b\|^2 = (A_i' A_i)^{-1} A_i' b$.

2. Let $d = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} (A_1' A_1) & & \\ & \ddots & \\ & & (A_n' A_n) \end{bmatrix}^{-1} A' b$.

3. Let $\tau = \underset{t}{\operatorname{argmin}}(\|Adt - b\|^2) = \frac{b'Ad}{\|Ad\|^2} = \frac{b'Ad}{\|A_{img}d\|^2 + \|A_{reg}d\|^2}$.

4. Let $v = d\tau$.

This algorithm first chooses a descent direction d with local Gauss-Newton optimizations at Step (1a), and then determines the step length τ with a higher-level Gauss-Newton optimization of the energy function given direction d. The algorithm can be understood as the following procedure. Each particle (or voxel) tries to take the shortest path to minimize the global energy as much as possible (Step (1a)), without knowing other particles' movements. Because their movements change the energy function jointly, rather than independently, a global step length τ is used to coordinate their movement. This strategy is fully compatible with Thirion's Demons algorithm [9] when A is separable.

At each optimization iteration, we do not choose as the descent direction the steepest direction $A'b$, for the following reason. The steepest direction $A'b$ sometimes is dominated by large $A_i'b$ at some voxels (as illustrated in Fig. 1), suppressing other voxels from displacing. Theoretically this phenomenon will disappear as $A_i'b$ approaching zero. However, in practice, on a discrete grid, this cannot be perfectly achieved. We noticed this phenomenon when we initially tried to use the steepest descent optimization.

This local Gauss-Newton strategy converges slower than the Gauss-Newton strategy, but it takes much less computation at each iteration. As shown in Section 2.4, its computational complexity is linear. In our experiments, it achieved equal registration goodness as the Gauss-Newton algorithm in [8], but significantly reduced the registration time and memory usage.

Though this local Gauss-Newton strategy is motivated to solve the problem of diffusion-tensor image registration, its application is not limited to diffusion-tensor images. Generally, it restores the $O(n)$ complexity to the Demons algorithm for situations where the gradient matrix is not separable. It also allows the Demons algorithm to directly incorporate regularization without increasing computational complexity. Application to HARDI image registration, a computation intensive problem, is another interesting extension.

$$A = \begin{bmatrix} 3 & 1 & & \\ 1 & 3 & & \\ & & 1 & 0.3 \\ & & 0.3 & 1 \end{bmatrix}, b = \begin{bmatrix} 3 \\ 3 \\ 0.5 \\ 1 \end{bmatrix}$$

$$min \left\| A \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} - b \right\|^2, v_1, v_2 \in R^2$$

Steepest   Local Gauss-Newton
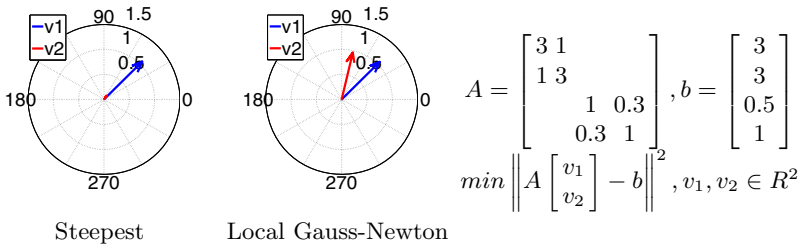
Fig. 1. Descent Directions. The steepest can be dominated by certain components.

2.4   Computational Complexity

The computational complexity of the algorithm is $O(n)$ where n is the number of voxels of interest. For a particular voxel, its tensor reorientation as well as its local energy function involves only the displacement of the voxels in its neighborhood, so A has a block sparse structure and the number and size of the non-zero blocks in A_i are constant numbers fully determined by image dimension and the regularizer's differential operator. Therefore, Step (1a) has $O(1)$ complexity, Step (3) has $O(n)$ complexity and the total complexity is $O(n)$.

3   Experiment

3.1   Data Acquisition and Preprocessing

Diffusion weighed images (DWI) of 120 pediatric subjects were collected with 30-direction isotropic DTI sequence (b = 1000s/mm^2, voxel size= 2 x 2 x 2 mm^3, dimension = 128 x 128 x 128). FSL brain extraction tool (BET) was applied to the B0 images to mask the brain region, with the "R" option turned. Due to the long computation time (about 7 hours) and high memory demand (about 20 GB) of the algorithm in [8], we randomly selected 12 pairs of the subjects for pair-wise registration.

3.2   Image Registration

For the diffeomorphic Demons algorithm proposed by Yeo *et al.* in [8], its open source implementation, the Tensor Toolkit (TTK) (`https://gforge.inria.fr/projects/ttk`) was used, with the parameters recommended in [8]: three level multi-scaling and each with ten iterations. For the Local Gauss-Newton Demons algorithm, we applied the affineness regularization to the displacement field. We adjusted the regularization weight to match the displacement smoothness of TTK's outputs. In Section 3.4, we evaluated the algorithms with three types of harmonic energy, include those that our algorithm did not directly optimize.

3.3   Evaluation Metric

Tensor rooted-mean-square (RMS) difference, FA correlation coefficients (Corr. Coef.), and MD Corr. Coef., between the fixed image and the transformed moving image, were used to indicate registration similarity. Three types of harmonic energy, affineness, curvature and diffusion, were used to indicate displacement smoothness. We used different harmonic energy to give a comprehensive evaluation, avoiding the bias introduced by algorithms' preference. All these metrics were estimated within the brain masks.

3.4   Results

For each of the registration cases, the two algorithms achieved similar matching between the fixed images and the warped images, with similar regularization energy, while the Local Gauss-Newton algorithm took significantly less computation time and memory usage, as shown in Table 1 and Figures 2 and 3. Regarding FA Corr. Coef, both algorithms performed almost equally well; regarding MD Corr. Coef and tensor RMS, the Local Gauss-Newton algorithm performed better, with smoother displacement fields.

Table 1. Summary Statistics of the Experiment

	Euclidean RMS	FA Corr. Coef.	MD Corr. Coef.	Time†
Local Gauss-Newton	0.700 ± 0.037	0.668 ± 0.027	0.707 ± 0.032	2272 ± 165
Gauss-Newton	0.741 ± 0.046	0.667 ± 0.033	0.656 ± 0.034	24441 ± 485
	Affineness	Curvature	Diffusion	Memory‡
Local Gauss-Newton	0.145 ± 0.008	0.108 ± 0.005	0.2680 ± 0.021	≈2 GB
Gauss-Newton	0.165 ± 0.006	0.119 ± 0.005	0.3056 ± 0.029	≈20 GB

†Computation time was estimated on an Intel Xeon 2.80 GHz CPU and single-threaded.
‡ Memory usage was manually monitored by using the Linux "top" command, during the algorithms worked on the finest resolution of the multi-scale registration.

This experiment with twelve cases suggests that the Local Gauss-Newton method can achieve similar goodness at similar harmonic energy as the method
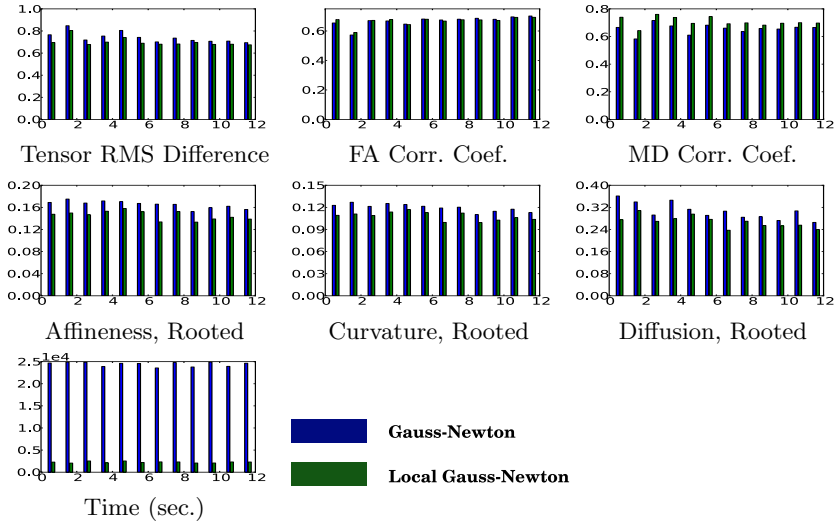
Fig. 2. Statistics of Each Registration Case. The two algorithms achieved similar goodness with similar regularization energy for each of the registration cases while the Local Gauss-Newton algorithm took much less computation time.

Fixed   Moving   Local Gauss-Newton   Gauss-Newton
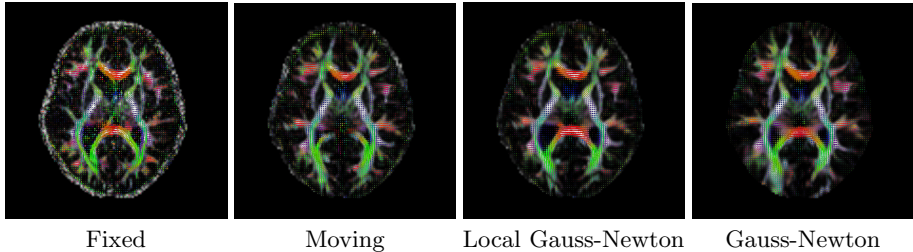
Fig. 3. Registration Results Example. Visually, the registration results of the Gauss-Newton and the Local Gauss-Newton methods look similar. Since they optimize the same energy function, as expected, they yield similar results.

in [8] does. However, the Local Gauss-Newton algorithm significantly reduced the computation time from about 7 hours to about 45 minutes, and decreased memory demand from 20 GB to 2 GB. The smoother displacement fields of the Local Gauss-Newton algorithm was possibly achieved by directly incorporating the affineness regularization into its optimization of the velocity fields.

4   Conclusion

As more and more brain-connectivity studies recruit a large number of subjects, it is important to design registration methods that are not only theoretically

rigorous, but also computationally efficient, for processing large data sets. However, the requirement of reorienting diffusion tensors complicates the calculation of deforming force, and significantly slows computation.

In this paper, we propose a method for the Demons algorithm to exactly incorporate both reorientation and regularization into deforming velocity calculation, but without directly solving a large non-separable linear system. This method restores the $O(n)$ computational efficiency of the original Demons algorithm to diffusion images, without sacrificing registration goodness, in comparison with directly solving a large linear system at each iteration. In our experiments, the new algorithm reduced the computation time and memory demand for ten times.

References

1. The Human Connectome Project, http://www.humanconnectomeproject.org/
2. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE Transactions on Medical Imaging 20(11), 1131–1139 (2001)
3. Park, H.J., Kubicki, M., Shenton, M.E., Guimond, A., McCarley, R.W., Maier, S.E., Kikinis, R., Jolesz, F.A., Westin, C.F.: Spatial normalization of diffusion tensor MRI using multiple channels. Neuroimage 20(4), 1995–2009 (2003)
4. Yang, J., Shen, D., Davatzikos, C., Verma, R.: Diffusion Tensor Image Registration Using Tensor Geometry and Orientation Features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 905–913. Springer, Heidelberg (2008)
5. Alexander, D.C., Gee, J.C.: Elastic matching of diffusion tensor images. Computer Vision and Image Understanding 77(2), 233–250 (2000)
6. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Med. Image Anal. 10(5), 764–785 (2006)
7. Cao, Y., Miller, M., Mori, S., Winslow, R., Younes, L.: Diffeomorphic matching of diffusion tensor images. In: Conference on Computer Vision and Pattern Recognition Workshop, CVPRW 2006, p. 67 (June 2006)
8. Yeo, T., Vercauteren, T., Fillard, P., Peyrat, J.M., Pennec, X., Golland, P., Ayache, N., Clatz, O.: DT-REFinD: Diffusion tensor registration with exact finite-strain differential. IEEE Transactions on Medical Imaging 28(12), 1914–1928 (2009)
9. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. 2(3), 243–260 (1998)
10. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuroimage 45(1 suppl.), S61–S72 (2009)
11. Fischer, B., Modersitzki, J.: A unified approach to fast image registration and a new curvature based registration technique. Linear Algebra and its Applications 380, 107–124 (2004)

Registration of Brainstem Surfaces in Adolescent Idiopathic Scoliosis Using Discrete Ricci Flow

Minqi Zhang[1], Fang Li[1], Ying He[1], Shi Lin[2,5], Defeng Wang[3,5], and Lok Ming Lui[4]

[1] School of Computer Engineering, Nanyang Technological University, Singapore
[2] Shenzhen Institutes of Advanced Technology, Chinese Academy of Science, Shenzhen, China
[3] CUHK Shenzhen Research Institute, Shenzhen, China
[4] Department of Mathematics, CUHK, Hong Kong
[5] Dept. of Imaging and Interventional Radiology,
The Chinese University of Hong Kong, Hong Kong

Abstract. Adolescent Idiopathic Scoliosis (AIS) characterized by the 3D spine deformity affects about 4% schoolchildren worldwide. Several studies have demonstrated the malfunctioning of postural balance, proprioception, and equilibrium control in patients with AIS. Since these functions are closely related to structures in and around the brainstem, the morphometry of the brainstem surface is of utmost importance. In this paper, we propose an effective method to accurately compute the registration between brainstem surfaces. Four consistent features, which describe the global geometry of the brainstem, are automatically extracted to guide the surface registration. Using the discrete Ricci flow method, brainstem surfaces are parameterized conformally onto the quadrilaterally-faced hexahedron, through which the brainstem registration can be obtained. Our registration algorithm can guarantee the exact landmark correspondence between brainstem surfaces. With the obtained registration, a shape energy can be defined to measure the local shape difference between different brainstem surfaces. We have tested our algorithms on 30 real brainstem surfaces extracted from MRIs of 15 normal subjects and 15 AIS patients. Experimental results show the efficacy of the proposed algorithm to register brainstem surfaces, which matches landmark features consistently. The computed registration can be used for the morphometry of brainstems.

1 Introduction

Adolescent idiopathic scoliosis (AIS) is a three-dimensional structural deformity of the spine that occurs during adolescence, and the prevalence is about 4% of the adolescent population worldwide. Due to the lack of generally accepted scientific theory on the etiology of AIS, the treatment and prognosis of AIS are limited [2]. Many experiments and evidences point toward the nervous regulation of the postural balance in the idiopathic scoliosis patients. The brainstem and vestibular system are two key organs in the balance control system. Significant anomalies of the balance function, proprioception and oculomotor reflexes have been reported [3], and morphological difference has been found in the vestibular system [4]. However, there is no reported work on studying the morphological alterations in brain stems in the AIS subjects in an objective and quantitative way. Motivated by this, we are interested in developing mathematical models which facilitate the morphometry of brainstem surfaces in AIS and normal controls.

In order to perform shape analysis effectively, meaningful one-to-one correspondences between different brainstems must be obtained. Such a process is called *surface registration*. Landmark-matching based registration approaches have been commonly applied, in which landmark features were required to be consistently matched to guide the registration. Landmark-matching registration has shown to be effective in obtaining accurate point-wise correspondences between 3D medical data. On brainstem surfaces, there are no medical features defined by neuroscientists that can be used as a constraint to establish good correspondences. Therefore, in order to register brainstems effectively, meaningful landmark features that describe the global geometry of the brainstem surfaces must firstly be extracted. On brainstem surfaces, four consistent feature curves can be observed, which are basically ridges and valleys with high curvatures. It motivates us to extract out these features to guide for geometric matching brainstem registrations. With these obtained features, landmark-matching registrations can be computed to obtain an accurate point registration that matches geometry as much as possible.

In this paper, we introduce an algorithm to automatically register brainstem surfaces based on the discrete Ricci flow method. Four consistent features are firstly extracted automatically to guide the surface registration. These features effectively describe the global geometry of the brainstem surfaces. Using the discrete Ricci Flow method, brainstem surfaces are parameterized conformally onto the quadrilaterally-faced hexahedron, of which the extracted feature landmarks are mapped to the edges. Quadrilaterally-faced hexahedron is chosen as the parameter domain since the geometry of the brainstem surface is similar to a hexahedron. With that, the distortion under the parameterization can be minimized so that registration can be obtained accurately. Surface registrations between brainstems can then be obtained through the parameterization, which consistently match the feature landmarks. With the obtained registration, a shape energy can be defined to measure the local shape difference between different brainstem surfaces, which is useful to study shape variation between brainstems for the purpose of disease analysis. We tested our algorithm on real brainstem surfaces extracted from MRIs of 15 normal subjects and 15 AIS patients. Experimental results show the efficacy of the proposed algorithms to register and detect the local shape difference between different brainstem surfaces.

Our contributions are two-folded: first, we propose to delineate four salient features on brainstems that describe the global geometry; second, by mapping the four detected landmarks to the four side edges of a hexahedron, we propose an efficient algorithm to conformally parameterize the brainstem surfaces, which naturally induces the registration among brainstem surfaces. Our registration algorithm can guarantee the exact landmark feature correspondence.

2   Previous Work

Surface registration, which aims to find a meaningful 1-1 correspondence between different surfaces, has been studied extensively by different groups. Conformal surface registration is commonly used [8][9], which gives a parameterization minimizing the angular distortions. An advantage of this approach is that they preserve local geometry very well. Conformal structure can also measure non-isotropic deformation effectively.
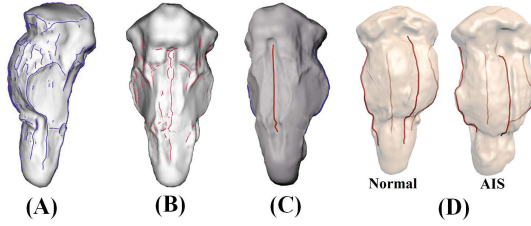
Fig. 1. Landmark extraction. (A) and (B) show the computed ridge and valley lines respectively. The two strongest ridge and valley lines are chosen as landmarks, as shown in (C). (D) shows the four consistent features extracted on brainstems of a normal and an AIS subject.

However, conformal registrations generally cannot map landmark features, such as sulcal landmarks on brain surfaces, consistently. Landmark-based diffeomorphisms are often used to compute, or adjust, cortical surface parameterizations [7,11]. For example, Glaunes et al. [7] proposed to generate large deformation diffeomorphisms of the sphere onto itself, given the displacements of a finite set of template landmarks. Leow et al. [11] proposed a level-set based approach to match different types of features, including points and 2D or 3D curves represented as implicit functions. These methods provide good registrations when the corresponding landmark points on the surfaces can be labeled in advance. On surfaces without well-defined landmarks, some authors have proposed driving features into correspondence based on shape information. Lyttelton et al. [13] computed surface parameterizations that match surface curvature. Fischl et al. [6] improved the alignment of cortical folding patterns by minimizing the mean squared difference between the average convexity across a set of subjects and that of the individual. Lord et al. [12] matched surfaces by minimizing the deviation from isometry.

3   Algorithm

3.1   Overview

Given the brainstem surface M, we first detect four salient geometric features on M. We then apply the discrete Ricci flow method to parameterize the brainstem surfaces onto a quadrilaterally-faced hexahedron, through which the four feature curves on the brainstem are mapped to the four vertical sides of the hexahedron. As our parameterization method naturally segments the brainstem surfaces into six patches, we can easily register two different brainstems by finding a diffeomorphism between each corresponding patches. In the following, we explain each step in details.

3.2   Landmark Extraction

On brainstem surfaces, four consistent features can be extracted, which are essentially curves with high surface curvatures. We extract these features by computing the ridge-valley lines [1], an effective shape descriptor on the surface along which the surface
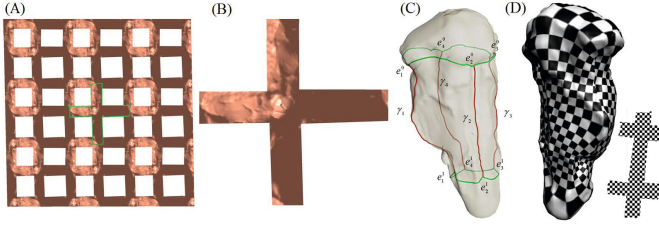
Fig. 2. The cut surface can be conformally embedded into the universal covering as shown in (A). (B) shows the fundamental domain (one period) of the universal covering.The parameterization naturally induces a segmentation of the brainstem surfaces into six patches as shown in (C). In (D), the checkerboard texture on the parameter domain is mapped to the brainstem surface, which mapping shows the parameterization is indeed conformal.

bends sharply. Figure 1(A) and (B) show the computed ridge-valley lines on a brainstem surface. After obtaining the ridge-valley lines, we smooth them using a Gaussian filter (with kernel size 2% of the main diagonal of the object's bounding box) to reduce the artifacts caused by the noise and/or poor triangulation. The detected ridges or valleys are merged, if their endpoints are close and their tangent directions are within the user-specified threshold. In our implementation, we set the distance and direction thresholds to 2% of the object's main diagonal and 15 degree respectively. Next, we measure the strength of each feature by computing the length of the detected ridge-valley line. The longer the curve, the stronger the feature. We then find the two strongest ridges and valleys, which are chosen as landmark features. On brainstem surfaces, there are consistent umbilic points and high-curvature points. These points are chosen as the endpoints of the landmark features. Fig. 1(C) shows the strongest ridge and valley lines extracted.

Among the four landmarks, we observed the two feature lines in the valley γ_2 and γ_4 are quite robust and consistent, since their end points fall in regions with highly negative mean curvatures. For the other two ridges γ_1 and γ_3, their locations are consistent, but their end points may differ among the subjects. To solve this issue, we specify the length of the ridges so that the tracing stops immediately when the landmark exceeds the threshold. We also allow the users to specify the end points manually. Fig. 1(D) shows the four consistent features extracted on brainstems of a normal and an AIS subject.

3.3 Parameterization Using Discrete Ricci Flow

To ease the computation process, we parameterize the brainstem onto a domain in $\mathbb{R}^2$. The ideal parameter domain for such surface is the quadrilaterally-faced hexahedron, of which each landmark is mapped to an edge of the hexahedron. Discrete Ricci flow method is used to obtain the parameterization.

Given a brainstem surface, we first cut it open along the detected feature lines. As a result, the open brainstem surface is of genus 0 with 4 boundaries, denoted by γ_i, $i = 1, ..., 4$. We then set the target curvatures to be $-\pi/2$ at the end points of the landmarks and zero elsewhere. Using the discrete Ricci flow method [10], we can parameterize the brainstem surfaces onto the universal covering space embedded in $\mathbb{R}^2$.
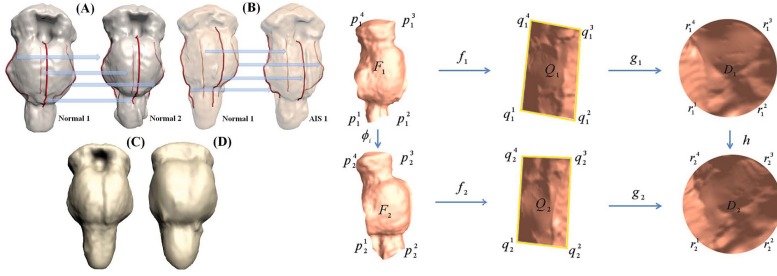
Fig. 3. Left: (A) Registration between two normal brainstems. (B) Registration between the brainstems of a normal and AIS subject. Landmark features are matched exactly. (C) shows the average shape of 15 AIS brainstems using our registration. The sharp features are well-preserved after averaging, meaning that our registration matches salient features well. (D) shows the average shape using the curvature-based sphereical demons registration without landmarks [15]. Note that the sharp features are smoothed out. Right:Computing the map $\phi_i : F_1^i \to F_2^i$.

To visualize the parameterization, we cut the brainstem surface along the geodesics (i.e. line segments) $e_1^1 e_2^1$, $e_2^1 e_3^1$, and $e_3^1 e_4^1$ (see Fig. 2(C)), under the new metric $\overline{g}$. Intuitively speaking, we cut along the edges of the bottom face to make the hexahedron open. Now the brainstem surface becomes a genus-0 surface with only one boundary.

We compute the layout of the parameterized mesh by isometrically embedding all the triangles with the new metric $\overline{g}$. Figure 2 shows the universal covering space of the brainstem surface with the highlighted fundamental domain. One can see that each landmark is mapped to the straight side of the hexahedron. This feature allows us to easily compute the high quality registration with guaranteed exact landmark matching, which will be described in the next subsection.

3.4   Registration

The above discrete Ricci flow method parameterizes each brainstem surface to a hexahedron, such that the four landmarks are mapped to the four vertical edges. The hexahedron naturally induces a segmentation of 6 patches. With these, we can easily register two brainstem surfaces in a piecewise manner.

Let F_1^i (resp. F_2^i), $i = 1, \cdots, 6$, denote the six patches of M_1 (resp. M_2) induced by the parameterization. We want to find a diffeomorphism (i.e., a C^∞-smooth and bijective map) $\phi_i : F_1^i \to F_2^i$ between each pair of patches F_1^i and F_2^i. Let f_1 (resp. f_2) denote the Ricci flow parameterization that maps F_1^i (resp. F_2^i) to a quadrilateral D_1^i (resp. D_2^i). Let us also denote the corners of the 3D patch and 2D quadrilateral by p^j and q^j, $j = 1, \cdots, 4$, respectively. Then we map each quadrilateral Q to a unit disc D by a harmonic function $g : Q \to D$ such that $\triangle g = 0$ with Dirichlet boundary condition $g(\partial Q) = \partial D = \mathbb{S}^1$. Let $r^j = f(q^j)$, $j = 1, \cdots, 4$, denote the images of the corners on the unit circle. Next, we compute another harmonic function $h : D_1^i \to D_2^i$ between the two unit discs D_1^i and D_2^i, i.e., $\triangle h = 0$. We set the boundary condition $h(\partial D_1^i) = \partial D_2^i$ by enforcing the exact feature correspondence, i.e., the corners r_1^j

are mapped to corners r_2^j. Therefore, a map ϕ_i between the patches F_1^i and F_2^i can be obtained by the composite map $\phi_i = f_2^{-1} \circ g_2^{-1} \circ h \circ g_1 \circ f_1$ (See Fig. 3 right).

A homeomorphism $\phi : M_1 \to M_2$ can then be obtained by $\phi = \bigcup_{i=1}^6 \phi$. Note that landmarks are exactly matched for each patches. So, our algorithm guarantees the exact correspondence between the landmark features.

Proposition. The proposed brainstem surface registration algorithm produces a homeomorphism ϕ between M_1 and M_2, which guarantees the exact correspondence between the landmarks $\phi(\gamma_1^j) = \gamma_2^j, j = 1, \cdots , 4$.

Proof. We first show that the resulting map ϕ is a homeomorphism. Observe that the discrete Ricci flow induced parameterization f_1 and f_2 are diffeomorphism [5]. According to the classical result [14], a harmonic function, which maps a $\Omega \subset \mathbb{R}^2$ to some convex region Ω' and maps the the boundary $\partial\Omega$ homeomorphically into the boundary $\partial\Omega'$, is diffeomorphic. As Q_j and $D_j, j = 1, 2$, are convex and the boundary conditions in g_j and h are homeomorphic, all the harmonic functions g_1, g_2 and h are diffeomorphic. Therefore, the map $\phi_i = f_2^{-1} \circ g_2^{-1} \circ h \circ g_1 \circ f_1$ is a diffeomorphism (see Fig. 3) right.

Although the maps ϕ_i's are calculated individually, they can be glued seamlessly. We use the arc-length parameterized boundary condition in computing the harmonic map from each patch to the unit disk. Assume two patches F_a and F_b share a boundary γ, and $p \in \gamma$ is an arbitrary point on γ. When we compute the map g from F_a to the unit disk, we set the Dirichlet boundary condition so that p's image is given by the arc-length parameterization. Similarly, the map from F_b to the unit disk also sends p to exactly the same location on the circle. With this, our method can guarantee the boundaries of two adjacent patches are mapped consistently and thus all patches are glued seamlessly in a C^0 manner. Thus, the global map, $\phi = \bigcup_i^6 \phi_i$, is in fact a homeomorphism.

Second, we show that the map ϕ preserves the correspondence between the landmarks. By setting the prescribed geodesic curvature on γ_1^j and γ_2^j to zero, $j = 1, \cdots , 4$, the metric $g(\infty)$ obtained by discrete Ricci flow can guarantee each landmark is mapped to a *line segment* (see Fig. 2(b)). Then with harmonic functions, g_1 and g_2, each line segment is mapped to an arc of the unit circle. Note that the exact feature correspondence are set when computing the harmonic function $h : D_1 \to D_2$ between two unit disks. Therefore, the composite function ϕ_i maps the boundary ∂F_1 homeomorphically to ∂F_2 and also sends each corner of F_1 to the corresponding corner of F_2. Thus, the map ϕ_i sends the landmark of F_1 to the corresponding landmark of F_2. As a result, the map ϕ preserves the exact correspondence between the landmarks $\phi(\gamma_1^j) = \gamma_2^j$, $j = 1, \cdots , 4$.

Putting it together, the map $\phi : M_1 \to M_2$ is a homeomorphism with guaranteed landmark correspondence. $\square$

3.5   Shape Variation Detection

With the obtained registration between brainstems, we can define a shape energy which detect local shape variations. According to Riemannian geometry theories, the local geometry of a Riemann surface can be described by its mean curvature and Gaussian curvature. Let B_1 and B_2 be two brainstem surfaces. Suppose $f : B_1 \to B_2$

Input: Two brainstem surfaces M_1 and M_2
Output: The homeomorphism $\phi : M_1 \rightarrow M_2$ with guaranteed landmark correspondence
1. Extract the four landmarks from the brainstem surfaces;
2. Cut the brainstem surfaces open along the detected landmarks;
3. Set the prescribed Gaussian and geodesic curvatures to each vertex and run Ricci flow;
4. With the resulting metric, embed the brainstem surfaces into quadrilaterally-faced hexahedral, which naturally induces a segmentation of M_1 and M_2 into six patches.
5. For each pair of patches $F_1^i \in M_1$ and $F_2^i \in M_2$, find the bijective map $\phi_i : F_1^i \rightarrow F_2^i$ by computing harmonic functions with Dirichlet boundary conditions;
6. Output the map $\phi = \bigcup_{i=1}^{6} \phi_i$.

Algorithm 1. Brainstem surfaces registration with guaranteed landmark correspondence.
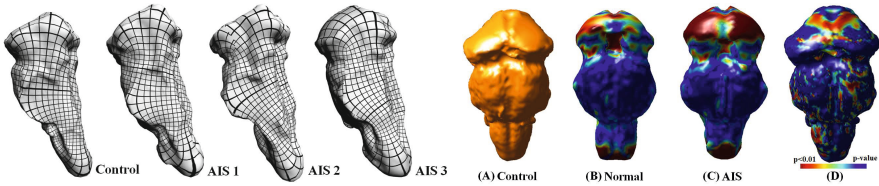
Control   AIS 1   AIS 2   AIS 3   (A) Control   (B) Normal   (C) AIS   (D)

Fig. 4. (A)-(D): Grid texture is drawn on the control brainstem. It is mapped to three different brainstems of the AIS subjects using our proposed registration algorithm. The grid pattern illustrates that our registration result is indeed an homeomorphism. (E)-(H):Detection of shape variations.

is the registration between B_1 and B_2. We can define a shape energy E_{shape} as follows: $E_{shape}(f) = \alpha \int_{B_1} |H_1 - H_2(f))| + \beta \int_{B_1} |K_1 - K_2(f))|$, where H_1 and K_1 (resp. H_2 and K_2) are the mean curvature and Gaussian curvature of B_1 (resp. B_2). $E_{shape}(f) = 0$ if and only if B_1 and B_2 are equal up to a rigid motion. Therefore, E_{shape} measures the local shape difference between B_1 and B_2 effectively. This allows us to examine the region of significant shape difference for the morphometry of the brainstems.

4   Experimental Results

Subject and data acquisition: We tested our proposed algorithms on 30 brainstem surfaces, which are extracted from the MRIs. The MRI brain data were acquired from 15 AIS subjects and 15 age-matched normal controls using a 1.5T MRI scanner (Sonata, Siemens, Erlanger, Germany) using a quadrature head-coil. The segmentation of the brain stems was achieved automatically. Two experienced operators verified the segmentation results.

Brainstem registration: Using our proposed algorithms, we compute the registration between 30 brainstems of 15 normal subjects and 15 AIS patients. Experimental results show that our proposed method can compute the registration effectively with exact landmark-matching. Figure 3(A) shows the registration results between two normal

brainstems. The registration result between the brainstems of a normal and AIS subject is shown in (B). Landmark features are exactly matched. In Fig. 4(A)-(D), grid texture is drawn on the control brainstem. It is mapped to three different brainstems of the AIS subjects using our proposed registration algorithm. Note that grid patterns have no over-lapping, indicating the registration is indeed a homeomorphism. The whole registration procedure can be computed efficiently. The feature landmarks (ridge and valley lines) are computed in real-time for brainstem meshes with 20K vertices. The Ricci flow based conformal parameterization for each brainstem mesh takes about 12 seconds, since the Newton's method with second order convergence rate is adopted in our implementation. The registration between brainstems through the parameterization takes about 5 seconds to compute. The whole process to obtain the landmark-matching brainstem registration takes less than 20 seconds. Hence, our proposed registration algorithm is quite efficient.

Detection of shape variations: With the obtained registration f, the local shape difference between different brainstem surface can be detected from the shape energy $E_{shape}(f)$. Figure 4(E)-(H) shows the local shape differences of two brainstem (1 normal and 1 AIS subject) surfaces from the control. The color map is given by the shape energy, which measures their local shape difference. The red (resp. blue) color indicates a higher (resp. lower) degree of shape difference. The local shape difference of the AIS brainstem from the control tends to be more obvious than that of the normal brainstem. Figure 4(H) shows the statistical significance p-map measuring the local shape difference from the control between the normal ($n = 15$) and AIS ($n = 15$) groups, plotted on a control brainstem. The deep red color highlights regions of significant statistical difference. This method can be potentially used to study factors that influence shape changes of brainstems in AIS.

Comparison: We also compare our registration algorithm with the curvature-based spherical demons registration without landmarks [15]. The curvature-based registration without landmarks generally cannot match salient features. Figure 3(C) shows the average shape of 15 AIS brainstems using our registration. The sharp features are well-preserved after averaging, meaning that our registration matches salient features well. (D) shows the average shape using the curvature-based registration without landmarks. Note that the sharp features are smoothed out. It indicates the mismatching of salient features.

5   Conclusion and Future Work

We present a rigorous algorithm to register brainstem surfaces for the disease analysis of Adolescent Idiopathic Scoliosis. The basic idea is to extract four consistent features, which describe the global geometry of the brainstem, to guide the surface registration. Using the Ricci Flow method, brainstem surfaces are parameterized conformally onto quadrilaterally-faced hexahedron, which naturally induces the feature landmark-matching brainstem registration. Our registration algorithm is a registration between brainstem surfaces with exact matching of the extracted feature curves. A shape energy can then be defined to measure the local shape variations between different brainstem surfaces. Experiments on real brainstem surfaces of 15 normal subjects and 15 AIS patients show the efficacy of our proposed algorithms to register and detect local shape

differences between different brainstem surfaces. The proposed algorithm can potentially be used to study factors that influence shape changes of brainstems in AIS. For future work, we will apply our proposed algorithms on more brainstem data to further investigate the relationship between the postural balance control problem and the occurrence of AIS. We would also like to point out that the proposed algorithm can be naturally extended for registration of other anatomical structures, such as hippocampus, cerebellum and the vestibular system, which will be investigated in our future work.

References

1. Belyaev, A.G., Anoshkina, E.V., Kunil, T.L.: Ridges: ravines, and singularities. Topological Modeling for Visualization 13, 375–383 (1997)
2. Miller, N.H.: Cause and natural history of adolescent idiopathic scoliosis. OCNA 30, 343–352 (1999)
3. Sahlstrand, T., Lidström, J.: Equilibrium factors as predictors of the prognosis in adolescent idiopathic scoliosis. Clin. Orthop. Relat. Res. 152, 232–245 (1980)
4. Zeng, W., Lui, L.M., Shi, L., Wang, D., Chu, W.C.W., Cheng, J.C.Y., Hua, J., Yau, S.-T., Gu, X.: Shape Analysis of Vestibular Systems in Adolescent Idiopathic Scoliosis Using Geodesic Spectra. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 538–546. Springer, Heidelberg (2010)
5. Chow, B., Luo, F.: Combinatorial Ricci flows on surfaces. JDG 63(1), 97–129 (2003)
6. Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284 (1999)
7. Glaunès, J., Vaillant, M., Miller, M.I.: Landmark matching via large deformation diffeomorphisms on the sphere. J. Maths. Imaging and Vision 20, 179–200 (2004)
8. Gu, X., Wang, Y., Chan, T.F., Thompson, P.M., Yau, S.-T.: Genus zero surface conformal mapping and its application to brain surface mapping. TMI 23(8), 949–958 (2004)
9. Hurdal, M.K., Stephenson, K.: Discrete conformal methods for cortical brain flattening. NeuroImage 45, 86–98 (2009)
10. Jin, M., Kim, J., Luo, F., Gu, X.: Discrete Surface Ricci Flow. TVCG 14, 1030–1043 (2008)
11. Leow, A., Yu, C.L., Lee, S.J., Huang, S.C., Protas, H., Nicolson, R., Hayashi, K.M., Toga, A.W., Thompson, P.M.: Brain structural mapping using a novel hybrid implicit/explicit framework based on the level-set method. NeuroImage 24(3), 910–927 (2005)
12. Lord, N.A., Ho, J., Vemuri, B.C., Eisenschenk, S.: Simultaneous registration and parcellation of bilateral hippocampal surface pairs for local asymmetry quantification. TMI 26, 471–478 (2007)
13. Lyttelton, O., Boucher, M., Robbins, S., Evans, A.: An unbiased iterative group registration template for cortical surface analysis. NeuroImage 34, 1535–1544 (2007)
14. Rado, T.: Aufgabe Jahresbericht der Deutschen Mathematiker-Vereinigung 35, 41–49 (1926)
15. Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration. TMI 29, 650–668 (2010)

Groupwise Rigid Registration of Wrist Bones

Martijn van de Giessen[1,2,3,4], Frans M. Vos[1,5],
Cornelis A. Grimbergen[4], Lucas J. van Vliet[1], and Geert J. Streekstra[4]

[1] Quantitative Imaging Group, Delft University of Technology, The Netherlands
[2] Division of Image Processing, Leiden University Medical Center, The Netherlands
[3] Department of Intelligent Systems, Delft University of Technology, The Netherlands
[4] Dept. of Biomed. Engineering and Physics, AMC Amsterdam, The Netherlands
[5] Dept. of Radiology, AMC Amsterdam, The Netherlands
m.vandegiessen@lumc.nl

Abstract. We present an extension of the symmetric ICP algorithm that is unbiased for an arbitrary number ($N \geq 2$) of shapes, using rigid transformations and scaling. The method does not require the selection of a reference shape or registration order and hence it is unbiased towards any of the registered shapes. The functional to be minimized is non-linear in the transformation parameters and thus computationally complex. We therefore propose a first order approximation that estimates the transformation parameters in a closed form, with computational complexity $\mathcal{O}(N^2)$.

Using a set of wrist bones, we show that the least-squares minimization and the proposed approximation converge to the same solution. Experiments also show that the proposed algorithms lead to smaller registration errors than algorithms that select a reference shape or register to an evolving mean shape. The low computational cost and trivial parallelization enable the alignment of large numbers of bones.

1 Introduction

Groupwise registration of multiple shapes (i.e. more than two) is a recurring problem in a wide variety of medical applications. Although much interest has been paid to non-rigid groupwise registration, especially intensity based such as in atlas building, some applications, e.g. the groupwise alignment of bone surfaces in orthopedics would benefit more from an unbiased *rigid* alignment. Although such an alignment can be obtained using the correspondence obtained with a non-rigid method, non-rigid registration results in general heavily depend on the careful tuning of the regularization parameters.

Many methods for the alignment of multiple shapes select one target to which all other shapes are registered or deformed, e.g. [7]. This, however biases the registration result to the selected shape. To minimize this bias, [5] proposed a strategy to select a shape that lies the closest the 'mean' shape, while others, e.g. [3,2] proposed to evolve a mean shape. The first method still does not completely remove the bias, while the latter methods add an extra layer of complexity, i.e. estimating the correspondence between registered example shapes and the mean shape.
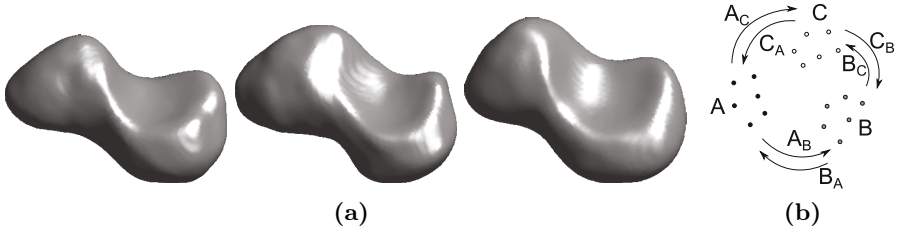
Fig. 1. (a) Examples of scaphoid bones from three individuals. (b) Three shapes with correspondence relations.

In this paper we propose a new method for the rigid registration (including isotropic scaling) of more than two objects that is inherently unbiased, i.e. it does not depend on the selection of a target shape or the order in which the shapes are processed. Moreover, the algorithm is stable and has a low computational complexity. This method is an extension of the symmetric ICP algorithm [1] to register more than two shapes. The direct extension yields a functional that is nonlinear in its transformation parameters and minimization is computational complex. To amend this we additionally introduce a method that involves separate closed-form estimations of the transformation parameters for each shape. We prove that this method converges to the same minimum as the direct ICP extension. Both algorithm variations are experimentally validated on a large number of subsets of 50 scaphoid bones, a bone in the wrist (See Figure 1a). These experiments demonstrate that the proposed algorithm leads to improved registration results over selecting the 'best' target shape as well as over registering to an evolving mean shape. Moreover, we will also show that the closed-form based variant converges to the same result as the direct extension of the extended ICP algorithm.

2   Methods

2.1   Unbiased ICP Algorithm for N Objects

We propose to extend the symmetric ICP algorithm to an unbiased algorithm for N shapes. For the sake of clarity we will initially present this algorithm for three shapes S_A, S_B and S_C, represented by point clouds. However, it can be directly generalised to N shapes with different representations. We define a set of control points for each shape, denoted by A, B and C. For each set of control points, the corresponding points on the other two shapes are determined using the minimum Euclidean distance as the criterion. For example, the corresponding points from S_B and S_C to A are B_A and C_A. A schematic representation of three shapes and their correspondences is shown in Figure 1b.

The transformation between point sets is found by minimizing the average squared Euclidean distance between corresponding point pairs. We extend the symmetric ICP to include all directional pairwise correspondences:

$$J_{nicp}(T_A, T_B, T_C) = \frac{1}{n_A} \|T_A(A) - T_B(B_A)\|^2 + \frac{1}{n_A} \|T_A(A) - T_C(C_A)\|^2$$
$$+ \frac{1}{n_B} \|T_B(B) - T_A(A_B)\|^2 + \frac{1}{n_B} \|T_B(B) - T_C(C_B)\|^2$$
$$+ \frac{1}{n_C} \|T_C(C) - T_A(A_C)\|^2 + \frac{1}{n_C} \|T_C(C) - T_B(B_C)\|^2 \quad (1)$$

in which T_A, T_B and T_C are transformations from point set A, B or C. This cost function is nonlinear in its transformations, since it contains products of the transformation parameters. Therefore, a closed-form solution of the transformations does not exist. The minimization can be done using a nonlinear regression technique, e.g. by the Levenberg-Marquardt algorithm.

Just as in the symmetric ICP algorithm, the minimization of (1) is mathematically ill-posed. The constraint $T_A = T_B^{-1}$, however, cannot be generalised to more than two transformations. Therefore, we choose the following constraints, that follow a similar rationale:

1. A net scaling of one: $\hat{s}_A \hat{s}_B \hat{s}_C = 1$
2. A zero net translation: $\hat{\mathbf{t}}_A + \hat{\mathbf{t}}_B + \hat{\mathbf{t}}_C = \mathbf{0}$
3. A zero net rotation: $\hat{\mathbf{r}}_A + \hat{\mathbf{r}}_B + \hat{\mathbf{r}}_C = \mathbf{0}$

in which $\hat{s}_A, \hat{s}_B, \hat{s}_C$ are the scalings, $\hat{\mathbf{t}}_A, \hat{\mathbf{t}}_B, \hat{\mathbf{t}}_C$ the translations and $\hat{\mathbf{r}}_A, \hat{\mathbf{r}}_B, \hat{\mathbf{r}}_C$ the rotations involved in the transformations T_A, T_B, T_C. The rotations are represented by the vectors $\hat{\mathbf{r}}_A, \hat{\mathbf{r}}_B, \hat{\mathbf{r}}_C$ that contain the Rodrigues parameters: the vectors $\hat{\mathbf{r}}_A, \hat{\mathbf{r}}_B, \hat{\mathbf{r}}_C$ are oriented parallel to the rotation axes (also called helical axes) of A, B and C and have a length equal to the magnitude of the rotation angle. These three constraints are included in the nonlinear least squares optimization using Lagrange multipliers.

2.2   Closed Form Transformation Estimates

A direct minimization of (1) is computationally expensive, because its non-linear form prevents finding the transformations T_A, T_B and T_C using a closed-form solution. We assume that the concurrent transformation of all clouds can be *approximated* by separate rigid registrations of each cloud that position the clouds at time step k 'in the middle' of all three clouds at time step $k-1$. 'The middle' is then defined as the position and orientation in which the sum of the quadratic Euclidean distances between the points in a cloud and their corresponding points in *all* clouds (including the transformed cloud at time step $k-1$) in the set is minimal. Accordingly, we define the functional $J_{approx}(T_A, T_B, T_C) = J_A(T_A) + J_B(T_B) + J_C(T_C)$ with right-hand terms of the form:

$$J_A(T_A^{(k)}) = \frac{2}{n_A} \left\| T_A^{(k)}(A) - T_A^{(k-1)}(A) \right\|^2$$
$$+ \frac{1}{n_B} \left\| T_A^{(k)}(A_B) - T_B^{(k-1)}(B) \right\|^2 + \frac{1}{n_A} \left\| T_A^{(k)}(A) - T_B^{(k-1)}(B_A) \right\|^2$$
$$+ \frac{1}{n_C} \left\| T_A^{(k)}(A_C) - T_C^{(k-1)}(C) \right\|^2 + \frac{1}{n_A} \left\| T_A^{(k)}(A) - T_C^{(k-1)}(C_A) \right\|^2 (2)$$

It may be observed that each of the functionals $J_A(T_A^{(k)})$, $J_B(T_B^{(k)})$ and $J_C(T_C^{(k)})$ depends on a single transformation, for which a closed-form solution exists. The first term in (2) effectively imposes that all shapes are equally involved. In comparison to (1), one transformation maps one shape unidirectionally onto the other shapes within each functional. The reverse mapping, i.e. mapping the other shapes onto the one, is actually contained in the other functionals.

Enforcing the constraints in Section 2.1 would result in a nonlinear problem. Instead, we opted to add a separate 'normalization' step after T_A, T_B and T_C are all updated.

For N shapes, N functionals as in (2) are needed, each with $\binom{N}{2}$ terms.

Normalization of scalings and rotations. After each update of all transformations, the net scaling of all shapes combined is $s_A s_B s_C = \bar{s}$, where s_A, s_B, s_C represent the current scale estimates. We adopt the next normalization to retain the ratio of the estimates:

$$\hat{s}_A = s_A/s_n, \hat{s}_B = s_B/s_n, \hat{s}_C = s_C/s_n \tag{3}$$

in which $s_n = \sqrt[3]{\bar{s}}$, such that $\hat{s}_A \hat{s}_B \hat{s}_C = 1$. The rotations undergo a similar normalization, so that the relative rotations between the three shapes is kept constant, i.e. $\hat{\mathbf{r}}_A + \hat{\mathbf{r}}_B + \hat{\mathbf{r}}_C = \mathbf{0}$. Let the sum of the Rodrigues parameters after minimizing (2) be $\mathbf{r}_A + \mathbf{r}_B + \mathbf{r}_C = \mathbf{r}_n$. Then rotations are normalized as follows

$$\hat{\mathbf{r}}_A = \mathbf{r}_A - \tfrac{1}{3}\mathbf{r}_n, \hat{\mathbf{r}}_B = \mathbf{r}_B - \tfrac{1}{3}\mathbf{r}_n, \hat{\mathbf{r}}_C = \mathbf{r}_C - \tfrac{1}{3}\mathbf{r}_n \tag{4}$$

If the rotation matrices that describe the updated rotations are $\hat{R}_A$, $\hat{R}_B$ and $\hat{R}_C$, the translations $\mathbf{t}_A, \mathbf{t}_B$ and $\mathbf{t}_C$ are updated to

$$\hat{\mathbf{t}}_A = \bar{\mathbf{a}}_T - \hat{s}_A \hat{R}_A \bar{\mathbf{a}}_S, \hat{\mathbf{t}}_B = \bar{\mathbf{b}}_T - \hat{s}_B \hat{R}_B \bar{\mathbf{b}}_S, \hat{\mathbf{t}}_C = \bar{\mathbf{c}}_T - \hat{s}_C \hat{R}_C \bar{\mathbf{c}}_S \tag{5}$$

where $\bar{\mathbf{a}}_T, \bar{\mathbf{b}}_T, \bar{\mathbf{c}}_T$ and $\bar{\mathbf{a}}_S, \bar{\mathbf{b}}_S, \bar{\mathbf{c}}_S$ are the means of A_T, B_T, C_T and A_S, B_S, C_S, respectively. It may be noticed that the value of $J_{approx}(T_A, T_B, T_c)$ is modified only due to the scale normalization, i.e. the normalization of rotation and translation parameters has no effect if $s_n = 1$.

2.3   Equivalence of Both Solutions

One may notice that when the algorithm in Section 2.2 has converged, it holds for all transformations that $T_X^{(k)} = T_X^{(k-1)}$ for all clouds $X \in \{A, B, C\}$. Thus, comparing (1) and (2), it holds at convergence that $J_{approx} = 2J_{nicp}$. From this one can see that there is a set of transformations for which the global minimum of J_{nicp} can be obtained with J_{approx}. Furthermore, the same transformations that lead to a global minimum of J_{nicp}, give a global minimum of J_{approx}: if the global minimum of J_{nicp} would not be the global minimum of J_{approx}, then it would be possible to find a set of transformations for which holds $J_{approx} < 2J_{nicp}$. The latter is not possible, because for any choice of $T_X^{(k)} = T_X^{(k-1)}$ it holds that $J_{approx} = 2J_{nicp}$. It can be proven that J_{approx} decreases every iteration similar to the pairwise ICP algorithm [1] and thus converges to a local minimum.

Table 1. Average values and standard deviations of J for $n = 400$ points, normalized to the results of Algorithm 3 and mean Target Registration Errors (mm).

N	$\bar{J}_1$	$\bar{J}_2$	$\bar{J}_4$	mTRE$_1$	mTRE$_2$	mTRE$_4$
2	1.006 (0.004)	1.288 (0.233)	1.000 (0.001)	1.4 (0.4)	1.4 (0.4)	1.4 (0.4)
4	1.069 (0.046)	1.544 (0.108)	1.005 (0.006)	1.4 (0.3)	1.5 (0.3)	1.5 (0.3)
8	1.146 (0.037)	1.573 (0.079)	1.003 (0.006)	1.6 (0.3)	1.7 (0.2)	1.5 (0.2)
16	1.177 (0.024)	1.576 (0.041)	1.000 (0.000)	1.7 (0.3)	2.1 (0.3)	1.4 (0.1)

3   Experiments

The performance of two existing algorithms (1) selecting the optimal mean shape [5] and (2) evolving a mean shape [1,6] and the proposed algorithms (3) in Section 2.1 and (4) in Section 2.2 are assessed in terms of accuracy and precision, as well as computational cost. These criteria are assessed as a function of the number of shapes, N, and the number of sampling points, n. CT images of 50 scaphoid wrist bones served as an application (See Figure 1a). Each scaphoid was represented by a point cloud, uniformly sampled on the bone surface. Algorithm 1 selects each of the N shapes as a target, registers all other shapes to this target and selects the result with the smallest remaining registration cost. Algorithm 2 takes one shape as initial mean and registers all shapes to this 'mean'. Subsequently the mean is updated by averaging the coordinates of corresponding points.

3.1   Accuracy and Precision

A subset of N bones was selected from the total of 50 and registered by Algorithms 1 to 4. For each outcome we evaluated J_{nicp} (1) in order to have comparable measures (effectively the algorithm from Section 2.1, Algorithm 3, serves as the reference standard). The resulting measures are J_1, J_2, J_3 and J_4. Each obtained value was divided by J_3 for normalization. The experiments were repeated 10 times for different subset selections from the 50 objects and for $N \in \{2, 4, 8, 16\}$ shapes and $n \in \{200, 400, 600, 800\}$ surface points. A maximum of $N = 16$ was taken since the reference Algorithm 3 became impractically time consuming for higher values. For the same reason a maximum of $n = 400$ points was used to register $N = 16$ shapes. Each of the N shapes involved in an experiment served as the target shape in Algorithm 2. The accuracies reported for Algorithm 2 were obtained both by averaging over all target shapes and subsets.

The mean normalized values $\bar{J}_{1,...,4}$ of $J_{1,...,4}$ for $n = 400$ and their standard deviations are shown in Table 1. $\bar{J}_4$ always differs less than 1% from $\bar{J}_3$. As expected, registering to a single target shape $\bar{J}_1$, for $N > 2$, leads to higher values than $\bar{J}_3$. Moreover $\bar{J}_1$ increases with increasing N. This follows directly from 1 as the registration in Algorithm 1 only has a (relatively decreasing) subset of the minimized terms in 1. The initial registration to a target shape in Algorithm 2 is by definition worse than Algorithm 1 as no selection of a shape closest to the mean is involved. Surprisingly, we found that after the initial
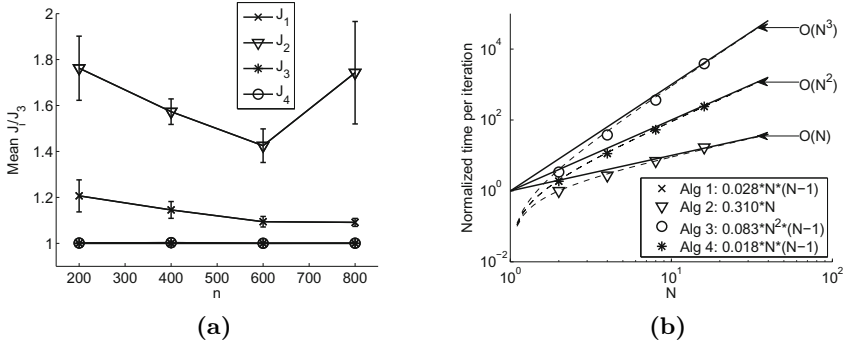
Fig. 2. (a) Registration results $\bar{J}_1$, $\bar{J}_2$, $\bar{J}_3$ and $\bar{J}_4$ as a function of the number of sample points for $N = 8$ shapes. Error bars denote standard deviations. Results of $\bar{J}_3$ and $\bar{J}_4$ almost coincide. (b) Normalized average processing times per iteration for Algorithms 1 ($\times$), 2 ($\triangledown$), 3 ($\circ$) and 4 ($*$) as a function of N for $n = 400$. The legend shows the functions fitted to the measured times per iteration. Plotted times are normalized to the numbers in the legend. Dashed: normalized theoretical times per iteration. Continuous: order of computational complexity. Results for Algorithm 1 ($\times$) and 4 ($*$) almost coincide.

registration to a target shape, $\bar{J}_2$ increased as the evolving mean was computed. This indicates that the evolving mean differs more from the reference mean than the selected target shape. The large standard deviations for Algorithm 2 are due to variable, non-optimal target shape selections. In all experiments J_2 was the highest, J_1 the second highest and J_3 or J_4 the lowest. The target shape that gave the lowest value of J in Algorithm 1 typically, but not always, yielded the lowest value of J in Algorithm 2. Registration differences between methods were about 5 degrees with extremes to 15 degrees. An evaluation using non-fiducial corresponding landmarks gave mean Target Registration Errors (mTRE) as in Table 1. Only for Algorithm 4 was the mTRE constant for increasing N.

Figure 2a shows that for an increasing number of points $\bar{J}_1$ decreases, while $\bar{J}_4$ has a negligible difference from $\bar{J}_3$. $\bar{J}_2$ seems to decrease too, for increasing n, but makes a sudden jump for $n = 800$. Although the standard deviation of $\bar{J}_2$ is large for $n = 800$, this result was not due to outliers and we attribute these results to convergence to local minima.

3.2 Computational Complexity

To assess the computational complexity of Algorithms 1 to 4, the execution time and the numbers of iterations were measured. The algorithms were implemented in MATLAB 7.4.0 on an AMD Opteron Quad Core Processor 250 at 2.0 GHz with 64 GB of memory, using a single core, to assess the computational complexity. The registration times were measered as the times needed until convergence within 1% of the final outcome to reduce the influence of the stopping criterium. Table 2 collates the mean registration times and their standard deviations for

Table 2. Average times in seconds needed for convergence of the four registration algorithms for $n = 400$. The standard deviations are between parentheses. For Algorithm 2, both the times for the initial registration to a target shape (I) and the registration to the evolving mean (II) are given.

N	Reg. 1	Reg. 2 (I)/(II)	Reg. 3	Reg. 4
2	1.3 (0.4)	0.6 (0.2)/0.6 (0.1)	7.0 (5.2)	0.8 (0.3)
4	7.13 (0.9)	1.8 (0.2)/2.4 (0.5)	93.7 (30.6)	5.5 (1.5)
8	28.0 (4.3)	3.5 (0.5)/7.6 (0.5)	1202.5 (147.2)	33.8 (8.8)
16	142.1 (1.2)	8.9 (0.1)/16.3 (3.6)	11560.5 (1304.0)	159.4 (18.8)

$n = 400$ surface points. Algorithm 2 is clearly the fastest, while Algorithms 1 and 4 take approximately the same time to converge. Algorithm 3 is impractically slow for large numbers of shapes. Running the registration on three cores (to leave one of the core available for the operating system), the registration of 8 and 16 shapes with Algorithm 4 took 11.9 and 55.5 seconds on average, a speedup of approximately 2.85. This shows that the algorithm can be efficiently parallelized. During the optimization, J_4 differed less than 1% from J_3 at each iteration, which is a strong indication that the computational scheme closely approximates the direct minimization both far from and close to the solution.

To estimate computational complexity as a function of the number of shapes N, the functions $a_1 \cdot N(N-1)$, $a_2 \cdot N$, $a_3 \cdot N^2(N-1)$ and $a_4 \cdot N(N-1)$ are fitted to the processing times per iteration. These fit functions can be derived from identifying the most expensive steps in the four algorithms such as correspondence finding and (in Algorithm 3) estimating the Jacobian. All four fits are shown in the legend of Figure 2b. Although the processing times per iteration are comparable between Algorithms 1 and 4, Algorithm 4 takes slightly longer for larger numbers of shapes N because the number of iterations of Algorithm 4 increases when N increases. The average processing times per iteration as a function of the number of points n was $\mathcal{O}(n \log n)$ for all four algorithms, corresponding to the complexity of closest point search using a Delaunay triangularization.

4 Discussion

We proposed a novel approach to extend the symmetric ICP algorithm to the unbiased registration of $N \geq 3$ shapes. All estimated correspondences between all shapes have equal weight in the cost functions (1) and the registration result does not depend on the order of registration.

Experiments confirmed that the proposed approximate minimization J_{approx} and a direct minimizationg of J_{nicp} converge to the same minimum. Assuming Gaussian distributed point correspondences, just as with the symmetric ICP algorithm, the proposed groupwise version gives optimal registration results since (1) is the minimum variance estimator for Gaussian distributed coordinates.

When choosing the best registration to a single target (Algorithm 1), the mean quadratic distance between shapes after registration, assessed by J_{nicp} (1), were all higher than after a direct minimization of J_{nicp}, and increased with

the number of shapes N. The difference, however, decreased with the number of surface points n. Registering to an evolving mean shape (Algorithm 2) resulted in a much higher mean quadratic distance between corresponding points. While direct minimization of J_{nicp} is practically not feasible for large numbers of shapes due to the order of time complexity $\mathcal{O}(N^3 \times n^2)$, the proposed approximation has a time complexity of order $\mathcal{O}(N^2 \times n^2)$. Thus this proposed approximation combines optimal registration accuracy with an acceptable time complexity. Furthermore, the algorithm was shown to be easily parallelizable, allowing an additional reduction registration time.

For clarity of the discussion, the unbiased ICP algorithms in this work estimate corresponding points in the same, separate step as the original ICP algorithm. This can, however, easily be improved, e.g. with a weighted closest point average as in [4], affecting only correspondence estimates such as B_A and C_A.

Using the proposed unbiased ICP algorithm and the approximation presented in this paper, it is possible to perform an unbiased registration with a computational complexity that allows the registration of a large number of bones. The unbiased registration ensures that the estimated bone alignment does not depend on the selection of a target or registration order.

References

1. Brett, A.D., Taylor, C.J.: A method of automated landmark generation for automated 3D PDM construction. Image and Vision Computing 18(9), 739–748 (2000)
2. Chui, H., Rangarajan, A., Zhang, J., Leonard, C.M.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(2), 160–172 (2004)
3. Frangi, A.F., Rueckert, D., Duncan, J.S.: Three-dimensional cardiovascular image analysis. IEEE Transactions on Medical Imaging 21(9), 1005–1010 (2002)
4. Granger, S., Pennec, X.: Multi-scale EM-ICP: A Fast and Robust Approach for Surface Registration. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 418–432. Springer, Heidelberg (2002)
5. Park, H., Bland, P.H., Hero III, A.O., Meyer, C.R.: Least Biased Target Selection in Probabilistic Atlas Construction. In: Duncan, J., Gerig, G. (eds.) MICCAI 2005, Part II. LNCS, vol. 3750, pp. 419–426. Springer, Heidelberg (2005)
6. Vos, F.M., de Bruijn, P.W., Aubel, J.G.M., Streekstra, G.J., Maas, M., van Vliet, L.J., Vossepoel, A.M.: A statistical shape model without using landmarks. In: Proceedings 17th International Conference on Pattern Recognition, ICPR17, Cambridge, UK, vol. 3, pp. 714–717. IEEE Computer Society Press (2004)
7. Yang, Y., Bull, A., Rueckert, D., Hill, A.: 3D Statistical Shape Modeling of Long Bones. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 306–314. Springer, Heidelberg (2006)

Automated Diffeomorphic Registration of Anatomical Structures with Rigid Parts: Application to Dynamic Cervical MRI

Olivier Commowick[1], Nicolas Wiest-Daesslé[2], and Sylvain Prima[1]

[1] INRIA, INSERM, VisAGeS U746 Unit/Project, F-35042 Rennes, France
University of Rennes I-CNRS UMR 6074, F-35042 Rennes, France
[2] CHU, University Hospital of Rennes, F-35043 Rennes, France
Olivier.Commowick@inria.fr

Abstract. We propose an iterative two-step method to compute a diffeomorphic non-rigid transformation between images of anatomical structures with rigid parts, without any user intervention or prior knowledge on the image intensities. First we compute spatially sparse, locally optimal rigid transformations between the two images using a new block matching strategy and an efficient numerical optimiser (BOBYQA). Then we derive a dense, regularised velocity field based on these local transformations using matrix logarithms and M-smoothing. These two steps are iterated until convergence and the final diffeomorphic transformation is defined as the exponential of the accumulated velocity field. We show our algorithm to outperform the state-of-the-art log-domain diffeomorphic demons method on dynamic cervical MRI data.

1 Introduction

In medical image analysis, one is often confronted with the problem of registering anatomical structures containing both hard, rigid (typically, bones) and soft, non-rigid (most other tissues) parts. Such problems are met for instance when following-up spinal cord lesions in MRI for the diagnosis of multiple sclerosis [1], or when assessing cervical injuries using dynamic/kinematic MR imaging with positional changes [2]. Many methods have been developed for both *fully* global rigid registration and *fully* local non-rigid registration separately [3], but the literature on hybrid methods, allowing for adequate registration of the structures depending on the *stiffness* of their components, is still quite sparse.

The earliest work we know of is that of Little et al. [4], who showed how to incorporate rigid structures into a deformation field, using radial basis functions; this was later improved by others to make the field invertible and even diffeomorphic [5,6,7]. However, these methods require the user to specify which structures are rigid, which led to the development of semi-automated methods in which rigidity can be locally favoured/enforced through a regularisation term in the criterion to be minimised [8]. This idea was later improved to allow for this term to be adaptively tuned to the structures to register, through prior segmentation of the rigid parts or design of a stiffness map (typically computed from

the image intensities; e.g. bones have high intensities in CT) [9,10,11]. Instead of segmenting rigid parts, it was also proposed to define several *anchors*, to which is attached an unknown polyaffine transformation, which can be subsequently estimated using a modified EM-ICP algorithm [12].

In this paper, we propose an iterative two-step method to compute a diffeomorphic non-rigid transformation between images of structures with rigid parts, without any user intervention or prior knowledge on the image intensities (to compute rigid parts or anchors). First we compute spatially sparse, locally optimal rigid transformations between the two images by adopting a new (as opposed to classical, translation-based) block matching strategy, made possible by the use of an efficient numerical optimiser (BOBYQA) (Sec. 2.1). The rationale behind this original strategy is our hope to recover both large rotations and subvoxel displacements. Then we derive a dense, regularised velocity field based on these local transformations using matrix logarithms and M-smoothing (Sec. 2.2). The floating image is then resampled and the two steps are iterated until convergence; the final diffeomorphic transformation is defined as the exponential of the accumulated velocity field. We finally compare our algorithm with the state-of-the-art log-domain diffeomorphic demons method [13] on dynamic cervical and multiple sclerosis MR images (Sec. 3).

2 Material and Methods

To compute a diffeomorphism T between a reference image I and a floating image J, we iterate between two steps: computation of a sparse set of locally optimal rigid transformations using block matching between I and $J \circ T^l$ (Sec. 2.1) and computation of a dense velocity field δLT^l computed from these locally estimated transformations (Sec. 2.2). Given that the transformation T is initially set to the identity ($T^0 = Id$), and that the initial velocity field is set to $LT^0 = \log T^0 = 0$, the velocity field is then updated as $LT^{l+1} = LT^l + \delta LT^l$. This two-step algorithm stops at the iteration l when δLT^l is close to 0, and the final diffeomorphism is computed as $T = T^l = \exp(LT^l)$. The complete algorithm is outlined in Sec. 2.3. For the sake of clarity, we detail the two steps using the simpler notations I, J and δLT (Sec. 2.1 and 2.2).

2.1 Computing a Sparse Set of Locally Optimal Rigid Transformations

Classical block matching algorithm. In this approach, that we *do not* follow, one first defines a set of blocks in each image, before matching each block in the reference image I with the most similar block in the floating image J. Similarity is typically computed using a measure on the voxel intensities, such as the sum of squared differences or the squared correlation coefficient in monomodal problems, or the mutual information or the correlation ratio in multimodal problems. The most common approach to optimise the similarity measure (at least in medical image analysis) is to perform an exhaustive search of the block with the highest similarity in J, within a given neighbourhood of each block in I.

This strategy implicitly assumes that the local motion between the images can be well recovered by a discrete translation (i.e. defined on the discrete grid of the image). Subvoxel displacements and large rotations are thus likely to be missed. It is all the more true when registering piecewise rigid structures, because in this case there exists no single, global rigid movement, that could be corrected before non-rigid registration.

Modified block matching algorithm. Recent advances in nonlinear optimisation allow for testing another strategy. We first define a set of blocks in I (as in the standard strategy), and then we propose to directly compute the rigid transformation best superposing each of these blocks with J, using a similarity measure on the voxel intensities. As opposed to the standard, discrete translation-based strategy, the computation of the similarity measure for a given block in I and a given tested transformation implies resampling to build the block in J. Given that the solution space is no longer finite, this leads to a potentially much more computationally expensive algorithm.

We propose to use the recent BOBYQA algorithm [14] to implement this idea. In essence, BOBYQA is a derivative-free, trust-region method which uses successive approximations of the similarity measure by quadratic functions, whose maxima can be computed analytically. It is very similar to the classical NEWUOA algorithm except that bounds must be specified on the variables. We thus end up with a set of blockwise-estimated optimal rigid transformations between I and J. In practice, however, we do not estimate a transformation for the blocks in I having a low variance σ^2. The set of estimated transformations $(R_1, \ldots, R_m)$ is thus spatially sparse, due to these missing transformations, and also due to the resolution of the grid of blocks in I, which is different from that of I. In addition, we weigh each estimated transformation R_i with a weight w_i set equal to the similarity measure; here we use the squared correlation coefficient, to be insensitive to local intensity changes, thus $0 \leq w_i \leq 1$.

2.2   Estimating a Dense Velocity Field

The set of estimated transformations $(R_1, \ldots, R_m)$ is spatially sparse, but is also noisy and likely to contain outliers (due to the noise in the images to be registered and the potential errors in local registrations). How to estimate a dense ($n = \text{card}(I)$) and smooth velocity field δLT from $(R_1, \ldots, R_m)$? We propose to use the logarithms of these m transformations, defined in the space of 4×4 real matrices ($\mathcal{M}_4(\mathbb{R})$) restricted to those whose last row contains only zeros, and to estimate n intermediate matrices in the same space (that we name $\log S_1, \ldots, \log S_n$ by analogy) as the minimisers of a criterion C:

$$(\log S_1, \ldots, \log S_n) = \underset{\log S_1, \ldots, \log S_n}{\arg \min} \left[\sum_{i=1}^{n} \sum_{j \in V_i} w_j \rho(|| \log S_i - \log R_j ||^2) d(|v_i - v_j|^2) \right],$$

where:

- $\rho : \mathbb{R} \to \mathbb{R}^+$ is a robust error norm,
- $||.||$ is the Frobenius norm in $\mathcal{M}_4(\mathbb{R})$, and $|.|$ is the Euclidean norm in $\mathbb{R}^3$,
- v_j is the coordinate of the central voxel of the block where R_j was estimated,
- v_i is the coordinate of the voxel where S_i is to be computed,
- V_i is a neighbourhood around the position v_i; note that the sum over $j \in V_i$ must read: "the sum over all the points v_j where a rigid transformation R_j was estimated, and which are inside V_i",
- w_j is the weight defined in Sec. 2.1,
- $d(.) : \mathbb{R}^3 \to \mathbb{R}^+$ is a (spatial) error norm.

It must be clear that we *do not* estimate S_i and then its logarithm; we *do* estimate $\log S_i$ directly; we use this notation here only as a convention for the sake of simplicity. Solving this minimisation problem is known as *local M-smoothing* [15], due to the use of a robust error norm and of spatial neighbourhoods to design C; C can be minimised using gradient descent, where each transformation can be estimated independently of the others. Using a particular adaptive, data-dependent step size leads to an easy-to-interpret update formula for each $\log S_i$ [15]:

$$\log S_i^{k+1} = \frac{\sum_{j \in V_i} w_j \rho'(||\log S_i^k - \log R_j||^2) d(|v_i - v_j|^2) \log R_j}{\sum_{j \in V_i} w_j \rho'(||\log S_i^k - \log R_j||^2) d(|v_i - v_j|^2)}$$

It can be seen from this formula that ρ' acts as a *tonal* kernel, while d acts as a *spatial* kernel. After convergence, each finally estimated $\log S_i$ is a linear combination of the logarithms of the rigid transformations R_j; following Arsigny et al. [5], we define the final dense velocity field δLT as $\delta LT(v_i) = \log(S_i).v_i$, $\forall i = 1, \ldots, n$. In practice, we define ρ as the Welsch function, which leads to $\rho'(a^2) = \exp(-a^2/2\lambda^2)$, and we define d as $d(b^2) = \exp(-b^2/2\theta^2)$; this leads to two similar expressions for the two kernels (with different bandwidths). V_i is spherical with radius 2θ (to achieve an approximate 95% confidence interval for a Gaussian law). To initialise the gradient descent algorithm, S_i^0 is computed as the solution of the update formula by setting $\rho'(a^2) = 1$ (i.e. no tonal kernel).

2.3   Complete Algorithm

The final estimated transformation is a diffeomorphism [5]. We perform all the update calculations on the velocity field, whose exponential is required only once per iteration to resample the floating image. This is the same approximation as that done by Vercauteren et al., who showed experimentally that $\exp(LT^{l-1}) \circ \exp(\delta LT)$ could be approximated by $\exp(LT^{l-1} + \delta LT)$ for a small enough velocity field δLT [13].

2.4   Implementation Details

For the **block matching**: size of the blocks: 7 voxels; grid step size: 3 voxels; minimal intensity variance in the blocks: 1/4 of the maximum squared intensity;

Algorithm

1: Initialize T to identity: $T^0 \leftarrow Id = \exp(LT^0)$ and the velocity field to 0: $LT^0 \leftarrow 0$
2: **for** each pyramid level of the multiresolution scheme, **do**
3:     **repeat**
4:         Estimate local rigid transformations using block matching (Sec. 2.1):
            $R = (R_1, \ldots, R_m) \leftarrow$ block-matching$(I, J \circ T^{l-1})$
5:         Interpolate a dense velocity field using M-smoothing (Sec. 2.2):
            $\delta LT \leftarrow$ M-smooth(R)
6:         Increment the velocity field: $LT^l = LT^{l-1} + \delta LT$
7:         Regularise (elastic-like) the velocity field: $LT^l \leftarrow G_\nu * LT^l$
8:         Compute $T^l = \exp(LT^l)$ to resample J
9:     **until** δLT is sufficiently small

search radiuses within BOBYQA: 2 voxels (translation) and 5 degrees (rotation). For the **M-smoothing**: kernel bandwidths: $\lambda^2 = \text{med}_{j \neq h} || \log R_j - \log R_h||^2/2$ (tonal), $\theta = 4$ voxels (spatial); $\nu = 4$ voxels in the elastic-like regularisation. We use a 3-level multiresolution strategy, and the resampling of the floating image is done using trilinear interpolation. Run-time of the algorithm (dual core Xeon 3.0 GHz PC): about 6 min (vs 2 min for the log-domain diffeomorphic demons [13]).

3 Validation and Results

We propose to assess our algorithm quantitatively on ten patients with traumatic cervical cord injury, who got dynamic cervical MRI (T2-w, size $384 \times 384 \times 14$, voxel size $0.8 \times 0.8 \times 3$ mm^3) with two different positions each: either flexion/neutral, or extension/neutral [2]. For each patient, we manually defined landmarks on the cervical/thoracic vertebrae C1-C3-C6-T1 (and T4 when visible), the pontomedullary junction, and the gnathion (lower border of the mandible) on each of the two MRI. We considered the neutral position as the reference image in the extension/neutral setting, and the flexion as the reference image in the flexion/neutral setting. For a given patient, the registration accuracy was evaluated as the root mean square error (RMSE) computed over the homologous landmarks after registration using four different methods: global rigid registration (M_1) [16], log-domain diffeomorphic demons (M_2) [13], our algorithm (M_3), and M_2 initialised using M_3 (M_4); both M_2 and M_3 are initialised using M_1. We also assessed our algorithm visually on two patients with MS lesions in the spinal cord, and two other patients with tumours in the spinal cord, with two time points each (T1-w, size $256 \times 256 \times 64$, voxel size $1 \times 1 \times 1$ mm^3).

The box-and-whisker plot in Fig. 1 computed from the ten patients shows our algorithm (M_3) to significantly outperform both M_1 (paired t-test: $p = 3 \times 10^{-4}$) and M_2 (paired t-test: $p = 2 \times 10^{-3}$), with much smaller error and much smaller error dispersal. M_3 is also slightly better than M_4 (paired t-test: $p = 5 \times 10^{-2}$). This suggests that the log-domain diffeomorphic demons performs worse than our algorithm even when properly initialised (using M_3 instead of M_1). The results of M_1, M_2 and M_3 on one of the ten patients are shown in Fig. 2, and that of M_1 and M_3 are shown on one of the MS patients in Fig. 3.
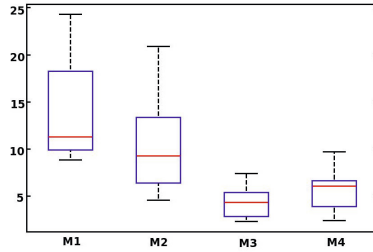
Fig. 1. Quantitative evaluation of registration accuracy. Box-and-whisker plot of registration errors (unit: millimetres) for the 4 compared methods (M_1 to M_4). The errors and the error dispersal are much smaller for M_3 compared to M_1 and M_2; M_4 is also slightly worse than M_3, which suggests that M_4 actually degrades the results compared to M_3 when initialised with M_3.
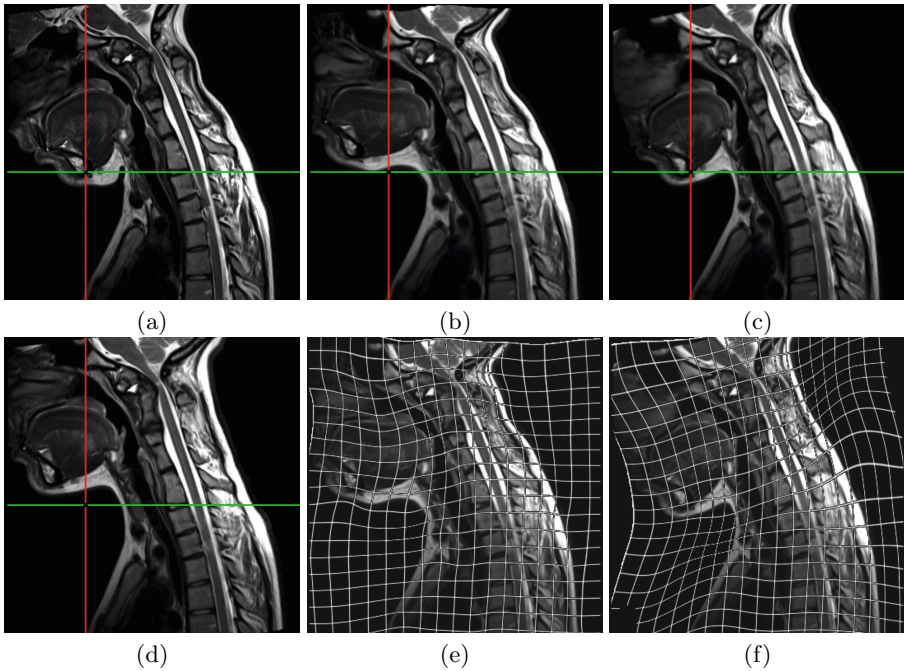
Fig. 2. Registration results on a patient with flexion/neutral positions. (a) reference image; (d,b,c) floating image registered to the reference image with M_1, M_2, M_3 respectively; (e,f) same as (b,c) with deformation grids overlaid. The intersection between the green and red lines shows the large error of M_2 on the mandible; on the contrary, M_3 correctly matches this point. The ability of M_3 to recover the flexion is further illustrated by the deformation grid: the deformation visually appears as near-rigid on the lower head and face, while it shows extension near the back of the neck and contraction near the front of the neck; on the contrary, the deformation grid shows that M_2 outputs near rigid movement everywhere.
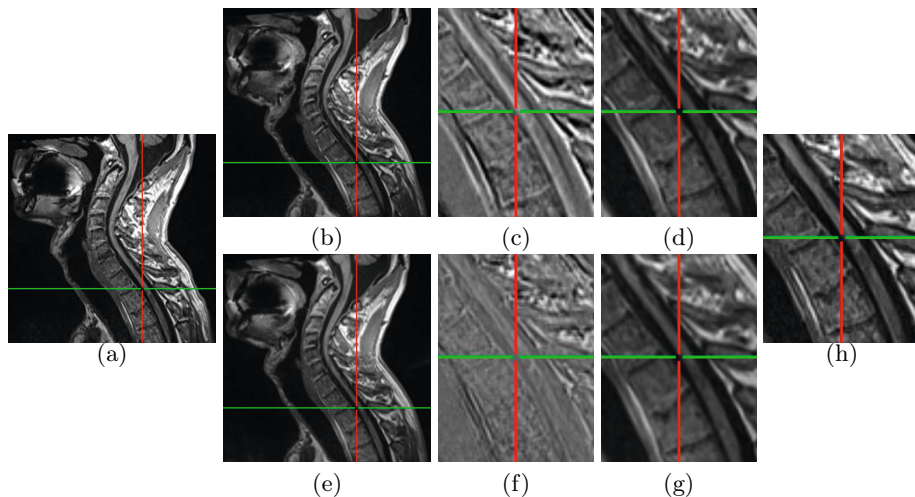
Fig. 3. Registration results on a patient with MS lesions in the spinal cord. (a) patient image at time point t_0; (b) patient image at time point t_1 registered with M_1 and (e) with M_3; (d,g) zoom on (b,e); (h) zoom on (a); (c,f) zoom on the difference (registered minus reference) image. Note that in this case M_2 (not displayed) performs as well as M_3. These snapshots, and in particular the zoomed difference images, visually show that M_3 gives a better result than a simple global rigid registration.

4   Conclusion and Perspectives

It appears that our strategy for non-rigid registration, based on the computation of locally optimal rigid transformations in the first place, allows us to recover displacements and deformations of piecewise rigid structures (as seen e.g. in dynamic cervical MRI) much better than standard methods which are implicitly based on locally optimal translations, such as the log-domain diffeomorphic demons algorithm. This original strategy was made possible by (i) the use of an up-to-date very efficient optimiser and (ii) the design of a specific regularising procedure on the (sparse) set of locally estimated rigid transformations, based on robust estimation techniques. A future line of research could be to combine our regularisation technique with those previously proposed in this context [9,10,11]. Our intuition is also that our algorithm could perform very well in more general problems, without necessarily rigid structures involved, and on other image modalities; we will evaluate this in a near future.

References

1. Okuda, D., Mowry, E., Cree, B., Crabtree, E., Goodin, D., Waubant, E., Pelletier, D.: Asymptomatic spinal cord lesions predict disease progression in radiologically isolated syndrome. Neurology **76**(8) (2011) 686–692
2. Abitbol, J.J., Hong, S.W., Khan, S., Wang, J.C.: Dynamic MRI of the spine. In Szpalski, M., Gunzburg, R., Rydevik, B.L., Huec, J.C., Mayer, H.M., eds.: Surgery for Low Back Pain. Springer (2010) 39–45
3. Maintz, J., Viergever, M.: A survey of medical image registration. Med Image Anal **2**(1) (1998) 1–36
4. Little, J., Hill, D., Hawkes, D.: Deformations Incorporating rigid structures. Comput Vis Image Und **66**(2) (1997) 223–232
5. Arsigny, V., Commowick, O., Ayache, N., Pennec, X.: A Fast and Log-Euclidean Polyaffine Framework for Locally Linear Registration. J Math Imaging Vis **33**(2) (2009) 222–238
6. Commowick, O., Arsigny, V., Isambert, A., Costa, J., Dhermain, F., Bidault, F., Bondiau, P.Y., Ayache, N., Malandain, G.: An efficient locally affine framework for the smooth registration of anatomical structures. Med Image Anal **12**(4) (2008) 427–441
7. Seiler, C., Pennec, X., Reyes, M.: Geometry-Aware Multiscale Image Registration via OBBTree-Based Polyaffine Log-Demons. In Fichtinger, G., Martel, A., Peters, T., eds.: MICCAI'2011. Vol. 6892 of LNCS, Springer (2011) 631–638
8. Rohlfing, T., Maurer, C.: Intensity-based non-rigid registration using adaptive multilevel free-form deformation with an incompressibility constraint. In Niessen, W., Viergever, M., eds.: MICCAI'2001. Vol. 2208 of LNCS, Springer (2001) 111–119
9. Loeckx, D., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using free-form deformations with a local rigidity constraint. In Barillot, C., Haynor, D., Hellier, P., eds.: MICCAI'2004. Vol. 3216 of LNCS, Springer (2004) 639–646
10. Ruan, D., Fessler, J.A., Roberson, M., Balter, J., Kessler, M.: Nonrigid registration using regularization that accomodates local tissue rigidity. In Reinhardt, J.M., Pluim, J.P., eds.: Medical Imaging 2006: Image Processing, SPIE Press (2006)
11. Staring, M., Klein, S., Pluim, J.P.: A rigidity penalty term for nonrigid registration. Med Phys **34**(11) (2007) 4098–4108
12. Taquet, M., Macq, B., Warfield, S.: Spatially Adaptive Log-Euclidean Polyaffine Registration Based on Sparse Matches. In Fichtinger, G., Martel, A., Peters, T., eds.: MICCAI'2011. Vol. 6892 of LNCS, Springer (2011) 590–597
13. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: A demons-based approach. In Metaxas, D., Axel, L., Fichtinger, G., Székely, G., eds.: MICCAI'2008. Vol. 5241 of LNCS, Springer (2008) 754–761
14. Powell, M.: The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report, Centre for Mathematical Sciences, University of Cambridge, UK (2009)
15. Mrazek, P., Weickert, J., Bruhn, A.: On robust estimation and smoothing with spatial and tonal kernels. In Klette, R., Kozera, R., Noakes, L., Weickert, J., eds.: Geometric Properties for Incomplete data. Springer (2006) 335–352
16. Ourselin, S., Roche, A., Prima, S., Ayache, N.: Block Matching: A General Framework to Improve Robustness of Rigid Registration of Medical Images. In Delp, S., DiGioia, A., Jaramaz, B., eds.: MICCAI'2000. Vol. 1935 of LNCS, Springer (2000) 557–566

Large Deformation Diffeomorphic Registration of Diffusion-Weighted Images

Pei Zhang[1], Marc Niethammer[2], Dinggang Shen[1], and Pew-Thian Yap[1,*]

[1] Department of Radiology,
[2] Department of Computer Science,
Biomedical Research Imaging Center (BRIC),
The University of North Carolina at Chapel Hill, USA
peizhang@email.unc.edu, mn@cs.unc.edu, {ptyap,dgshen}@med.unc.edu

Abstract. Registration of Diffusion-weighted imaging (DWI) data emerges as an important topic in magnetic resonance (MR) image analysis. As existing methods are often designed for specific diffusion models, it is difficult to fit to the registered data different models other than the one used for registration. In this paper we describe a diffeomorphic registration algorithm for DWI data in a large deformation setting. Our method generates spatially normalized DWI data and it is thus possible to fit various diffusion models *after* registration for comparison purposes. Our algorithm includes (1) a reorientation component, where each diffusion profile (DWI signal as a function on a unit sphere) is decomposed, reoriented and recomposed to form the orientation-corrected DWI profile, and (2) a large deformation diffeomorphic registration component to ensure one-to-one mapping in a large-structural-variation scenario. In addition our algorithm uses a geodesic shooting mechanism to avoid the huge computational resources that are needed to register high-dimensional vector-valued data. We also incorporate into our algorithm a multi-kernel strategy where anatomical structures at different scales are considered simultaneously during registration. We demonstrate the efficacy of our method using *in vivo* data.

1 Introduction

DWI registration presents a direct way of establishing correspondences for white matter micro-structures, which are often elusive in anatomical scans, such as T_1- and T_2-weighted images. As it is required to deal with both spatial alignment of macro-structures and reorientation of local angular structures, DWI registration is more challenging to develop than traditional scalar-based image registration.

DWI data are often acquired in up to hundreds of diffusion-sensitizing gradient directions so as to precisely delineate local angular structures. Various diffusion models are often fitted to the acquired data for analytical purposes. However, analysis can not yet be performed without aligning similar structures across different subjects. To this end, a number of registration algorithms have thus been developed. Geng et al. [1] used

a spherical harmonic (SH) representation of orientation distribution functions (ODFs) to guide registration. Yap et al. [2] developed a hierarchical registration scheme where the alignment is refined using features extracted from a SH-based representation with gradually increasing order. Raffelt et al. [3] utilized a subject-template-symmetric diffeomorphic framework to align fiber orientation distribution (FOD) fields. Hong et al. [4] performed registration with the help of T_2-weighted images and applied the resulting deformation fields to the diffusion-weighted images with re-transformation—taking into account rotation, scaling, and shearing effects of the spatial transformation of the FOD. Du et al. [5] designed a large deformation framework to register the ODF data.

However, all of the above methods are designed for specific diffusion models, which makes it difficult to fit other models to the registered data. In this paper we describe a method that is able to generate spatially normalized DWI data so that one can fit any diffusion model *after* registration. Key highlights of our method include:

1. **DWI Reorientation:** Our method can directly reorient DWI diffusion signal profiles.
2. **Large Deformation:** We use a large deformation diffeomorphic metric mapping (LDDMM) framework [6] to tackle large structural variations. Spatial image alignment is achieved by optimizing over a spatio-temporally varying velocity field.
3. **Geodesic Shooting:** A major problem with the LDDMM algorithm [6] is the large memory consumption. This is often aggravated for vector-valued and high dimensional DWI data. We thus use a geodesic shooting algorithm [7] to avoid the storage of the entire series of velocity fields, so that only an initial image and an initial momentum are needed to parameterize the full deformation path.
4. **Multi-Kernel:** We use multiple Gaussian kernels [8] to simultaneously register anatomical structures at different scales.

Works on registering the raw DWI data are few. To the best of our knowledge, the only closest work is that of Dhollander et al. [9], where they achieved the goal by using SHs [3] as well as a diffeomorphic demons algorithm [10]. Our method differs fundamentally from theirs in three aspects: (1) We achieve reorientation by using Watson distributions instead of SHs. This avoids the computational complexity of SHs as well as the loss of sharp directional information when SH basis functions of insufficient order are used; (2) Our method can work with single-shell DWI data, whereas their method requires multi-shell data acquisition, which might be clinically infeasible; (3) Our method explicitly considers large deformation.

2 Methodology

Below we will first describe the approach to reorienting DWI data in Q-space. We will then focus on the simplified shooting algorithm used for registration. Finally, a summary of the proposed method will be given.

2.1 Reorientation of DWI Data

To reorient the DWI data in Q-space we first decompose the diffusion signal profile into a series of fiber basis functions (FBFs), which are based on the Watson distribution function [11]. We then apply a local transformation, computed from the estimated
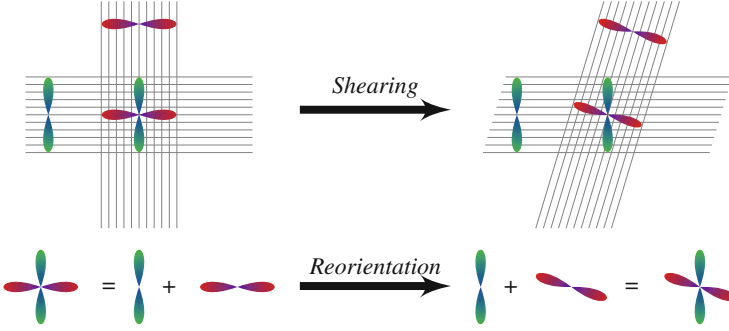
Fig. 1. Two fiber populations (gray lines) are shown together with their individual diffusion signal profiles. When the two fiber populations cross each other, the acquired diffusion signal profile is a combination of the responses from both fiber populations. As each fiber population transforms differently with respect to a local transformation (horizontal shearing in this example), the profile at the crossing should be decoupled, reoriented individually, and then recombined to form a reoriented diffusion signal profile.

map, to reorient each FBF independently. Finally, we recompose the reoriented FBFs to obtain the orientation-rectified DWI profile. See Fig. 1 for an illustration.

Diffusion Profile Modeling. Let $S(\mathbf{q}_i)$ be the diffusion signal measured in direction $\mathbf{q}_i$ $(i = 1, \ldots, M)$. Our goal is to represent $S(\mathbf{q}_i)$ in terms of a series of FBFs. As we use a set of Watson distributions [11] to realize the FBFs, we can write

$$S(\mathbf{q}_i) = w_0 f_0 + \sum_{j=1}^{N} w_j f(\mathbf{q}_i | \boldsymbol{\mu}_j, \kappa), \;\; \kappa < 0, \tag{1}$$

where $f(\mathbf{q}|\boldsymbol{\mu}, \kappa) = C(\kappa) \exp(\kappa(\boldsymbol{\mu}^{\mathrm{T}} \mathbf{q})^2)$ is the probability density function of the Watson distribution [11], $\mathbf{q}$ and $\boldsymbol{\mu}$ are unit vectors indicating the diffusion gradient direction and the mean orientation respectively, κ is a constant, and $C(\kappa)$ is the normalization factor. w_j is the weight associated with each FBF $f(\cdot)$. $f_0 \equiv C(0)$ is a constant representing the isotropic diffusion component. Given the diffusion signal profile $\mathbf{S} = [S(\mathbf{q}_1), \ldots, S(\mathbf{q}_M)]^{\mathrm{T}}$, we have $\mathbf{S} = \mathbf{Fw}$, where $\mathbf{w} = [w_0, w_1, \ldots, w_N]^{\mathrm{T}}$ and

$$\mathbf{F} = \begin{bmatrix} f_0 & f(\mathbf{q}_1 | \boldsymbol{\mu}_1, \kappa) & \cdots & f(\mathbf{q}_1 | \boldsymbol{\mu}_N, \kappa) \\ \vdots & \vdots & \ddots & \vdots \\ f_0 & f(\mathbf{q}_M | \boldsymbol{\mu}_1, \kappa) & \cdots & f(\mathbf{q}_M | \boldsymbol{\mu}_N, \kappa) \end{bmatrix}.$$

Since typically, $M < N + 1$, we have a set of underdetermined linear equations. We solve this using a L_1 regularized least-squares solver with a non-negative constraint. The reader is referred to [12] for details of the algorithm and evaluation.

Transformation and Recomposition. To reorient the direction of each FBF, $\boldsymbol{\mu}_j$, we apply a local affine transformation $\mathbf{A}$ estimated from the map resulting from registration, i.e. $\boldsymbol{\mu}'_j = \mathbf{A}\boldsymbol{\mu}_j / \|\mathbf{A}\boldsymbol{\mu}_j\|$. A matrix of rotated FBFs, $\mathbf{F}'$, can be then computed

based on $\boldsymbol{\mu}'_j$. The transformed DWI signal $\mathbf{S}'$ is finally computed as $\mathbf{S}' = \mathbf{F}'\mathbf{w}$. Note that the isotropic component is not rotated.

2.2   A Simplified Geodesic Shooting Algorithm

We now describe the registration method for a pair of DWI data. To avoid the computational complexity of a full adjoint shooting method [13], we follow the simplified shooting approach [7]. However, we modify it to allow for a gradient descent directly on the initial Hamiltonian momentum (the co-adjoint variable to the transported image, instead of the vector-valued momentum).

Below we use I to represent a vector-valued image of a diffusion profile $\mathbf{S}$ at each voxel location, and I^i to denote the i-th channel of I. Let I_0 be the source image and I_1 be the target image. Our goal is to minimize

$$E(v, I) = \frac{1}{2} \int_0^1 \|v\|_V^2 \, dt + \frac{1}{\sigma^2} \sum_{i=1}^M \|I^i(1) - I_1^i\|_{L^2}^2, \text{ s.t. } I_t^i + \nabla(I^i)^\mathrm{T} v = 0, \ I^i(0) = I_0^i,$$

(2)

where v is the sought-for time-dependent velocity field, $\sigma > 0$ is a constant and $\|v\|_V^2 = \langle L^\dagger L v, v \rangle_{L^2}$, where L is a proper differential operator. Instead of defining L we define a desired smoothing kernel $K = (L^\dagger L)^{-1}$. We use a multi-Gaussian kernel [8] to introduce a natural multi-resolution property to the solution and to provide an intuitive way of parameter tuning based on the desired scales that should be captured by the registration. Note that we run our algorithm with multiple iterations to minimize (2). In each iteration, I_0 is spatially transformed and reoriented using the map estimated in the previous iteration (see Sect. 2.3 for details).

The minimization of (2) leads to the following optimality and boundary conditions:

$$\begin{cases} I_t^i + \nabla(I^i)^\mathrm{T} v & = 0, \ I^i(0) = I_0^i, \\ -p_t^i - div(p^i v) & = 0, \ p^i(1) = -\frac{2}{\sigma^2}(I^i(1) - I_1^i), \\ L^\dagger L v + \sum_{i=1}^M p^i \nabla I^i & = 0. \end{cases}$$

(3)

Note that $\nabla_v E = L^\dagger L v + \sum_{i=1}^M p^i \nabla I^i$. Hence, instead of solving (3) as a boundary value problem [6] we follow a simplified shooting approach [7], performing the gradient descent only for $t = 0$. In contrast to [7], here we perform the gradient descent directly on the $\{p^i(0)\}$ by pulling the final conditions $\{p^i(1)\}$ back to $t = 0$. This can be accomplished by computing a backward map (from $t = 1$ to $t = 0$) on the fly during a forward integration (from $t = 0$ to $t = 1$). To obtain the gradient with respect to these momentum variables note that at convergence (or on a geodesic in general) for all times $L^\dagger L v + \sum_{i=1}^M p^i \nabla I^i = 0$. Therefore at $t = 0$

$$L^\dagger L \delta v(0) + \sum_{i=1}^M \delta p^i(0) \nabla I^i(0) = 0$$

(4)

because $I(0) = I_0$ is known.

In the vector-valued version of the standard LDDMM scheme [6] (which, given a velocity field v uses a forward sweep for I^i and a backward sweep for p^i) the Hilbert gradient at $t = 0$ is computed as

$$\nabla_{v(0)} E = v(0) + (L^\dagger L)^{-1} \left(\sum_{i=1}^{M} \hat{p}^i(0) \nabla I^i(0) \right),$$

where $\hat{p}^i(0)$ is the adjoint at $t = 0$ obtained after the forward sweep for I^i, which allows the computation of $\hat{p}^i(1)$ followed by a backward sweep for p^i. Since $v(0) = -(L^\dagger L)^{-1}(\sum_{i=1}^{M} p^i(0) \nabla I^i(0))$ is the initial velocity given the current initial momentum $p^i(0)$ the gradient can be rewritten as

$$\nabla_{v(0)} E = (L^\dagger L)^{-1} \left[\sum_{i=1}^{M} \left(\hat{p}^i(0) - p^i(0) \right) \nabla I^i(0) \right].$$

Substituting into (4) we obtain

$$\sum_{i=1}^{M} \delta p^i(0) \nabla I^i(0) = \sum_{i=1}^{M} \left(p^i(0) - \hat{p}^i(0) \right) \nabla I^i(0).$$

Since this needs to hold for any initial image $I(0)$ it follows that $\nabla_{p^i(0)} E = p^i(0) - \hat{p}^i(0)$. Given the (on-the-fly) computed map Φ which maps $t = 1$ to $t = 0$ the gradient is then $\nabla_{p^i(0)} E = p^i(0) - |D\Phi| \hat{p}^i(1) \circ \Phi$.

2.3 Summary of the Approach

We first use the method described in Sect. 2.1 to decompose both I_0 and I_1, and then run the above geodesic-shooting LDDMM to iteratively transform and reorient I_0. Specifically, we first estimate a global affine transformation $\mathbf{A}_g$ between the anisotropy images of I_0 and I_1. And then at each iteration, we (1) reconstruct both I_0 and I_1 with a decreasing κ and an increasing number of diffusion directions. I_0 is reconstructed using the FBFs reoriented with the map estimated in the previous iteration together with $\mathbf{A}_g$, while I_1 is reconstructed using the original FBFs with an identity map; (2) weight each reconstructed image using the associated anisotropy image; (3) estimate the map between the weighted images; and (4) compose the resulting map with the one estimated in the previous iteration. At the end of the registration we will obtain the final map between I_0 and I_1 as well as a transformed and reoriented source image I_0'.

3 Experiments

The DWI data were acquired from 11 adults using a Siemens 3T TIM Trio MR Scanner with an EPI sequence. Diffusion gradients were applied in 120 non-collinear directions with diffusion weighting $b = 2000\,\text{s/mm}^2$. The imaging matrix is 128×128 with rectangular FOV of $256 \times 256\,\text{mm}^2$. 80 contiguous slices with a slice thickness of 2 mm cover the whole brain.

We randomly chose a subject and used the associated data as the target image, and used the rest of the data as the source images. We computed the anisotropy images of all data and used them to estimate a set of affine transformations $\{\mathbf{A}_g\}$. We then warped each source image to the target image using the associated $\mathbf{A}_g$ and computed the mean image of the anisotropy images of the warped source (Fig. 2b). This blurred mean image implies that the DWI data cannot be well aligned by affine transformation.

We then used our method (3 iterations) to register each source image to the target image. After registration, we reconstructed each source in the original 120 directions using the associated $\mathbf{A}_g$ together with the resulting map. Averaging the anisotropy images across the subjects leads to the mean image shown in Fig. 2d. Repeating the above process with the map generated in the first iteration gives the mean in Fig. 2c. We found that our method significantly outperforms affine registration by producing a much crisper mean. Further improvement can be achieved by running the registration multiple times.

To quantify the comparison, we computed the RMS error between the vector-valued voxels at corresponding positions. This was done between the target and each source image, warped either using $\{\mathbf{A}_g\}$ or the map estimated by our method. Averaging the resulting RMS error images across the subjects leads to the mean images shown in Fig. 2e-g. The mean and standard deviation (s.d.) of these mean images are 13.5±6.2 for affine registration, 11.0±4.8 for the first iteration, and 10.3±4.6 for the final iteration.
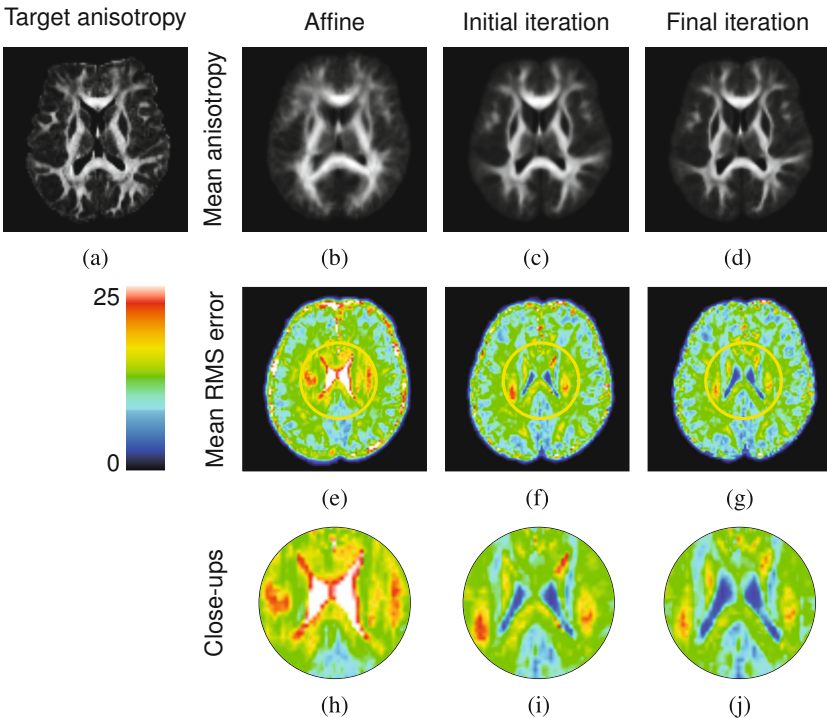
Fig. 2. Comparison of registration accuracy between affine registration and our method
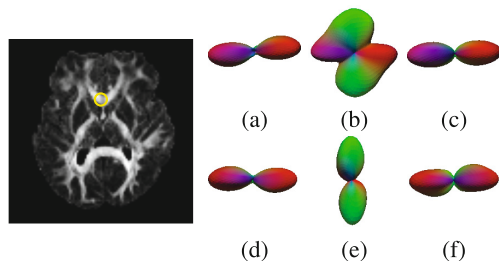
Fig. 3. A region of interest is indicated by the yellow circle in the left figure and some typical ODFs associated with this region are given on the right. (a) and (d): the ODFs of the target image; (b) and (e): the ODFs of the mean image generated using affine registration; (c) and (f): the ODFs of the mean image generated using our method.
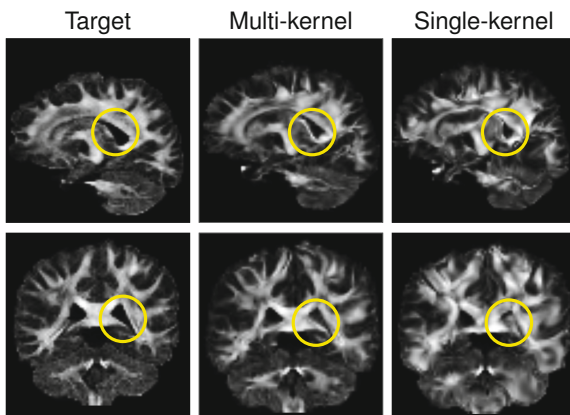
Fig. 4. Comparison of registration results given by multiple Gaussian kernels and a single kernel

We also labeled a set of salient landmark points (around the lateral ventricles) on the anisotropy images. For affine registration we used $\{\mathbf{A}_g\}$ to warp the landmarks of each source to the target space, and computed the mean and s.d. of the Euclidean distances (in mm) between the warped landmarks and the corresponding landmarks on the target. For our method this was done by using $\{\mathbf{A}_g\}$ as well as the resulting maps. The results for affine registration, the first iteration and the final iteration are 7.0 ± 3.0, 4.4 ± 2.4 and 3.6 ± 2.2.

The results from these two quantitative comparisons are in agreement with our observations based on the mean anisotropy images.

Figure 3 shows that at voxel level the alignment of the DWI profiles across subjects can greatly benefit from registration using the proposed method. The ODFs of the average DWI data can deviate significantly when registration is inaccurate.

To demonstrate the advantage of using multiple Gaussian kernels, we repeated the above experiment by using only one Gaussian kernel. Figure 4 clearly shows that a single kernel is unlikely to capture all shape variations that are present in the data.

4   Concluding Remarks

We have described a method for direct registration of DWI data. Our method is capable of producing spatially normalized DWI data, based on which any diffusion model can be fitted for comparison purposes. Future work includes unifying the registration and reorientation steps for further improvement on alignment accuracy as well as comparing the performance of different diffusion models within the proposed framework.

References

1. Geng, X., Ross, T.J., Zhan, W., Gu, H., Chao, Y.-P., Lin, C.-P., Christensen, G.E., Schuff, N., Yang, Y.: Diffusion MRI Registration Using Orientation Distribution Functions. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 626–637. Springer, Heidelberg (2009)
2. Yap, P.T., Chen, Y., An, H., Yang, Y., Gilmore, J.H., Lin, W., Shen, D.: SPHERE: Spherical harmonic elastic registration of HARDI data. NeuroImage 55(2), 545–556 (2011)
3. Raffelt, D., Tournier, J.D., Fripp, J., Crozier, S., Connelly, A., Salvado, O.: Symmetric diffeomorphic registration of fibre orientation distributions. NeuroImage 56(3), 1171–1180 (2011)
4. Hong, X., Arlinghaus, L., Anderson, A.: Spatial normalization of the fiber orientation distribution based on high angular resolution diffusion imaging data. Magnetic Resonance in Medicine 61(6), 1520–1527 (2009)
5. Du, J., Goh, A., Qiu, A.: Diffeomorphic metric mapping of high angular resolution diffusion imaging based on riemannian structure of orientation distribution functions. IEEE Transactions on Medical Imaging 31(5), 1021–1033 (2012)
6. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
7. Ashburner, J., Friston, K.J.: Diffeomorphic registration using geodesic shooting and Gauss Newton optimisation. NeuroImage 55(3), 954–967 (2011)
8. Risser, L., Vialard, F.X., Wolz, R., Murgasova, M., Holm, D.D., Rueckert, D.: Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. IEEE Transactions on Medical Imaging 30(10), 1746–1759 (2011)
9. Dhollander, T., Veraart, J., Van Hecke, W., Maes, F., Sunaert, S., Sijbers, J., Suetens, P.: Feasibility and Advantages of Diffusion Weighted Imaging Atlas Construction in Q-Space. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 166–173. Springer, Heidelberg (2011)
10. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)
11. Jupp, P., Mardia, K.: A unified view of the theory of directional statistics. International Statistical Review 57, 261–294 (1989)
12. Yap, P.T., Shen, D.: Spatial transformation of DWI data using non-negative sparse representation. IEEE Transactions on Medical Imaging (to appear)
13. Vialard, F.X., Risser, L., Rueckert, D., Cotter, C.: Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. International Journal of Computer Vision 97(2), 229–241 (2012)

Prediction of Brain MR Scans in Longitudinal Tumor Follow-Up Studies

Lior Weizman[1], Liat Ben-Sira[2], Leo Joskowicz[1], Orna Aizenstein[2],
Ben Shofty[2], Shlomi Constantini[2], and Dafna Ben-Bashat[2]

[1] School of Eng. and Computer Science, Hebrew University of Jerusalem, Israel
[2] Sourasky Medical Center, Tel-Aviv, Israel
`lweizm45@cs.huji.ac.il`

Abstract. We present a new method for the estimation of the next brain
MR scan in a longitudinal tumor follow-up study. Our method effectively
incorporates information of the past scans in the time series to predict
the future scan of the patient. Its advantages are that it requires no user
intervention and does not assume any particular tumor growth model.
Instead, the patient-specific tumor growth parameters are estimated in-
dividually from the past patient scans. To validate our method, we con-
ducted an experimental study on four patients with Optic Path Gliomas
(OPGs) and four patients with glioblastomas multiforma (GBM), each
scanned at five time points. The tumor volumes in the predicted and ac-
tual future scans, both segmented by an expert radiologist, yield a mean
volume overlap difference of 13.65% for the OPG patients, and 34.23%
for the GBM patients.

1 Introduction

Brain tumor detection, characterization, and follow-up using CT and MR images
is the current standard of care in neuroradiology. The decision making process
takes into consideration previous scans of the patient, and aims at detecting
visual clues (markers) that indicate tumor progression, regression, or stable dis-
ease. However, the estimation of these trends is a difficult task due to tumor
inhomogeneity, inter-scans variations, and the lack of practical tools to estimate
the tumor volume consistently and reliably.

The prediction of tumor growth is a challenging task. The natural growth
of tumor cells does not always adhere to a specific growth model. Additionally,
changes in the tumor growth resulting from different therapies are still largely
unknown and unexplored, and therefore are not amenable to analytic modeling.
Thus, in practice, prediction methods aim at providing a coarse estimate of the
future involvement areas of the tumor, rather than predicting the exact future
spatial extent and volume of the tumor.

The current literature on tumor progression prediction consists of two main
approaches. The first one relies on tumor growth models. The models describe
the evolution of the tumor cells over time at a microscopic and/or macroscopic
level. Hogea et al. [1] propose a model that couples glioma growth with the

subsequent deformation of the brain tissue. Clatz et al. [2] describe a model that simulates the volumetric growth of glioblastomas multiforma (GBM). The model is derived from two subsequent MRI scans of a patient acquired at six months intervals. While promising, the model parameters are independently estimated from a single scan without incorporating the other parameters related to the natural history of the patient.

The second approach to tumor progression prediction assumes a model derived from the tumor growth pattern from the patient scans. For example, [3] shows how to predict response to therapy based on the derived apparent diffusion coefficients (ADCs) values computed from diffusion-weighted magnetic resonance imaging (DWMRI). Loncaster et al. [4] use measured volumes and signal intensities in T1 MRI scans to predict radiotherapy outcome. Note that these works focus on the prediction of response to a specific treatment on a predefined area of interest rather than aim at predicting an entire future scan.

We present a new algorithm to compute the predicted MRI scan of the entire brain based on imaging markers trends derived from the past scans of the patient in a longitudinal study (the second approach above). The method computes the optical flow of the voxels over time, and uses past motion vectors to predict a new deformable field, which is then used to estimate the next scan in the time series. Two key advantages of our method are that it takes into account multiple past scans of the patient and that it can be applied to a variety of brain tumors types since the tumor growth model prediction parameters are estimated from the scans of the patient itself instead of from tumor-specific growth models.

Since most tumor types do not have predictable growth patterns, clinical implementation of the method would focus on providing a visual interpretation of future involvement regions of the tumor, under the assumption that it continues its trends from the past. This additional source of visual information can be very helpful in determining a situation of stable disease and for the decision making process of patients who undergo a long-term and non-invasive treatment.

Universally accepted gold standards for the validation of future predicted scan have not been developed yet. To provide an error measure for our method, we conducted a study in which experienced radiologists blindly manually segmented the tumor regions in the current, predicted, and actual future scans. The study included four patients with Optic Path Gliomas (OPGs) and four with GBMs. The comparison between the tumors segmented volumes in the predicted scans vs the actual future scans yield a mean surface distance error of 0.4mm and 0.83mm, and a mean volume overlap difference of 13.65% and 34.23% for the OPG and GBM patients, respectively. To the best of our knowledge, this is the first method that estimates a future scan based on patient specific tumor growth trends in previous scans.

2   Method

The inputs to our method are $N + 1$ consecutive rigidly registered (with the method in [5]) T1-weighted contrast-enhanced pulse sequence MRIs of the same patient acquired at roughly equal time intervals, $I_t, I_{t-1}, ..., I_{t-N}$. The output is the estimation of the next scan of the patient in the time series.

The method consists of two steps. In the first step, the optical flow over time for every voxel in the current scan is computed. In the second step, the future transformation fields are estimated with an autoregressive model for groups of voxels. Fig. 1 shows a representative slice from the past, current, predicted, and actual future scans.

2.1   Temporal Optical Flow Computation

In this step, we compute the spatial path of each voxel in the current scan over time. First, we find the non-rigid transformation field between every two consecutive scans of the patient. We use the method of Kroon et al. [6], itself adapted from [7], to MRI scans. For every pair of consecutive scans, the method computes three transformation fields, one for every axis. This results in N forward and backward transformation fields. The forward transformation fields are $\{\mathbf{F}_{t-i}(\bar{r})\}_{i=1}^{N}$, where $\mathbf{F}_{t-i}(\bar{r})$ is the 3-dimensional forward transformation field from scan I_{t-i} to scan I_{t-i+1}, and $\bar{r} = (x, y, z)$ represents the spatial location in the scan. Similarly, $\{\mathbf{B}_{t-i}(\bar{r})\}_{i=1}^{N}$ are the backward transformation fields.

Next, we define the motion matrix optical flow for the patient's scans. The matrix $\tilde{\mathbf{V}}$ is a 5-dimensional matrix whose dimensions are $(scan\ dimensions) \times 3 \times N$, representing the 3D motions of every voxel in the scan over time. We define $\mathbf{V} = \tilde{\mathbf{V}}(\bar{r}, \cdot, \cdot)$ as a $3 \times N$ matrix, where $\mathbf{V}(\bar{r}, i)$ is a 3×1 vector representing the motion of the voxel in location $\bar{r}$ from time $(t-i)$ to $(t-i+1)$. The elements of $\mathbf{V}(\bar{r}, i)$ are recursively computed by:

$$\mathbf{V}(\bar{r}, i) = F_{t-i}(\bar{r}_{t-i+1} + B_{t-i}(\bar{r}_{t-i+1})) \tag{1}$$

where

$$\bar{r}_{t-i+1} = \begin{cases} \bar{r} & \text{when } i = 1 \\ \bar{r}_{t-i+2} + B_{t-i+1}(\bar{r}_{t-i+2}) & \text{otherwise} \end{cases} \tag{2}$$

and $\bar{r}$ is the location of the voxel in the latest scan, I_t. $\mathbf{V}(\bar{r}, \cdot)$ is the spatial path over time of a voxel from the baseline scan I_{t-N} to the current scan I_t.

2.2   Estimation of Future Transformation Field

Next, we estimate the future motion of a voxel based on its past motions. To have sufficient statistical predictive power, we estimate the model parameters for every group of voxels assuming that voxels with similar past motion values follow the same prediction model. For this, we cluster all the voxels in the image into groups based on their past spatial flow.

We use the K-means clustering algorithm to group the voxels in the scan spatial space into K clusters based on their spatial flow over time, $\mathbf{V}$. We obtain K groups of voxels, where C_k is the number of voxels in the k-th group.

To predict the future scan of the patient, we estimate $\mathbf{V}(\bar{r}, t)$, i.e. we predict the future motion optical flow for each voxel. For every voxels cluster, we model the future motion as a linear combination of past motions, expressed by an autoregressive (AR) model:
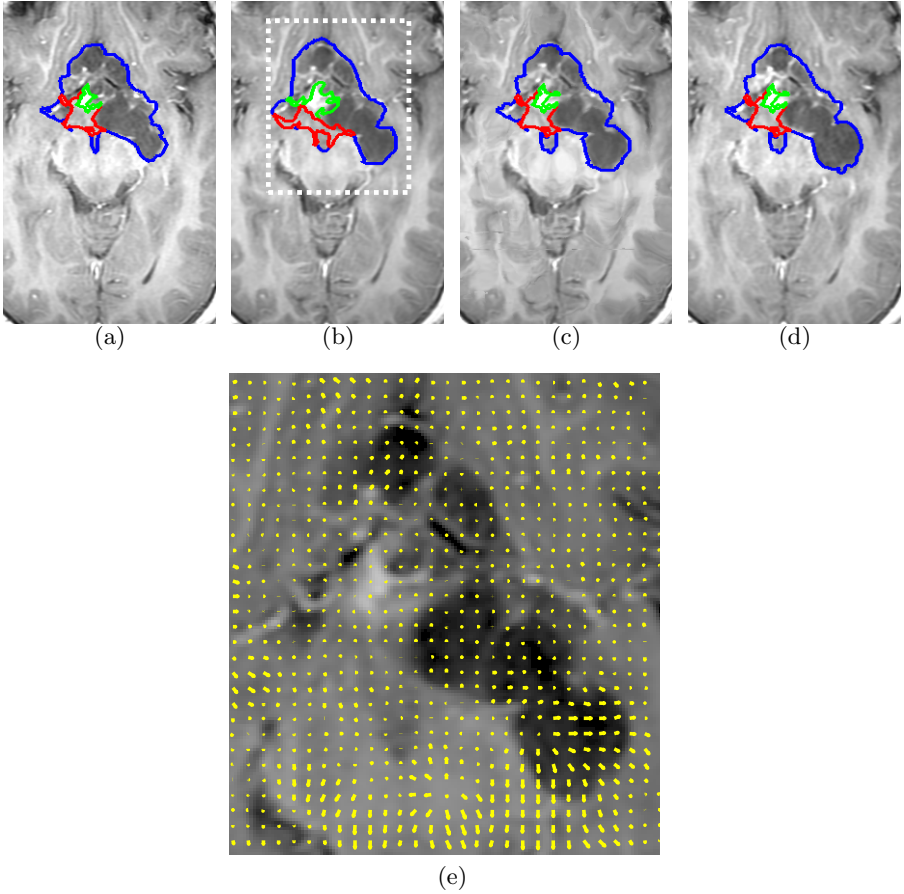
Fig. 1. Illustration of our method on a representative slice of T1-weighted pulse sequence of an OPG tumor and its internal components: (a) past scan; (b) current scan; (c) predicted scan; (d) actual acquired future scan – the tumor components are solid (red), enhancing (green) and cyst (blue), and; (e) estimated future transformation field projected on the scan plane and overlaid on the dashed area of the current scan – the yellow arrows indicate the local direction of the estimated future transformation field. Observe in the bottom right region of the image in (e) that the algorithm successfully predicted the expansion of the cyst component of the tumor.

$$\bar{v}_l(i) = \sum_{j=1}^{p} \mathbf{A}_k(j) \cdot \bar{v}_l(i - j) \tag{3}$$

where $\bar{v}_l(i)$ denotes the 3D optical flow motions from time $(t-i)$ to $(t-i+1)$ of a voxel in the k-th group, and p is the model order, $p \leq N$. Note that k is the index over clusters and l is the index of voxels in the cluster. The 3×3 matrices $\mathbf{A}_k(j)$ include the AR coefficients of the process, each corresponding to a specific lag, for the k-th cluster. This model describes a Multiple-Input-Multiple-Output (MIMO)

system, where the same prediction coefficients are used for all the voxels in the k-th group. Minimizing the forward prediction error in the least squares sense for this model, we obtain the Yule-Walker equations [8] of the MIMO model:

$$\sum_{j=1}^{p} \mathbf{A}_k(j)\mathbf{R}_v(i-j) = \mathbf{R}_v(i) \quad 1 \leq i \leq p \tag{4}$$

Note that, unlike the traditional MIMO case where $\mathbf{R}_v(n)$ is the autocorrelation of a multivariate process at lag n, here $\mathbf{R}_v(n)$ is the spatial mean of the temporal autocorrelation functions of the v_l at lag n, i.e.:

$$\mathbf{R}_v(n) = E_l(E_i[(\bar{v}_l(i)) \cdot (\bar{v}_l(i+n))]) \tag{5}$$

Assuming that $(v_1(n), v_2(n), ..., v_{C_k}(n))$ are independent, identically distributed (i.d.d) with zero mean for every n, the sample covariance is:

$$\hat{\mathbf{R}}_v(n) = \frac{1}{C_k} \sum_{l=1}^{C_k} \frac{1}{N-n} \sum_{j=1}^{N-n} \bar{v}_l(j) \cdot \bar{v}_l(n+j)^T \tag{6}$$

Substituting (6) in (4) yields a set of linear equations which are solved with the Levinson-Durbin recursion [8].

The result is an estimation for the forward transformation field, $\mathbf{F}_t(\bar{r}) = \mathbf{V}(\bar{r}, t)$. Figs. 1(e) and 2(e) show examples of these fields. Finally, this transformation field is applied to the current image, I_t to obtain the estimation of the next future scan in the time series, $\hat{I_{t+1}}$.

3   Experimental Results

We conducted a retrospective quantitative evaluation of our method with clinical gadolinium enhanced T1-weighted MRI scan series of eight patients. Patients did not undergo surgical interventions during the period of the research. Patients were divided into two groups. The first group consisted of four Neurofibromatosis (NF) patients, 3-7 years old subjects with OPGs which were scanned at approximately 6-month intervals. The second groups consisted of four adult GBM patients which were scanned at approximately 2-month intervals. All scans were acquired by a General Electric Signa 1.5T HDXT scanner. Each scan has $512 \times 512 \times 90$ voxels with isotropic voxel size of $1mm^3$.

Every patient was scanned five times at different time points. The first four scans were used to predict the fifth scan with our method. Based on empirical tests, we set the model order p to 3. We tested the method's performance for different values of K in the range 50-500 and we obtained the optimal results for our datasets at K=100. An expert radiologists manually produced tumor segmentations for the current, predicted, and the actual future scans using Analyze Direct 7.0 (Mayo Clinic, Rochester, MN).

Figs. 1 and 2 show representative results of the method. Table 1 shows the volume overlap error (VOE) and the mean surface distance (SD) of the manually

segmented tumor volumes in the predicted scans vs. the actual future scans. Detailed explanation about our validation measures can be found in [9]. For comparison, the same validation metrics were also computed for the segmented tumors in the current scan vs. the actual future scan. It can be seen that the method provides better results on GBM than OPG. This phenomenon can be explained by the fact the GBM is an aggressive tumor, that changes much more rapidly than OPG.

(a)   (b)   (c)   (d)

(e)
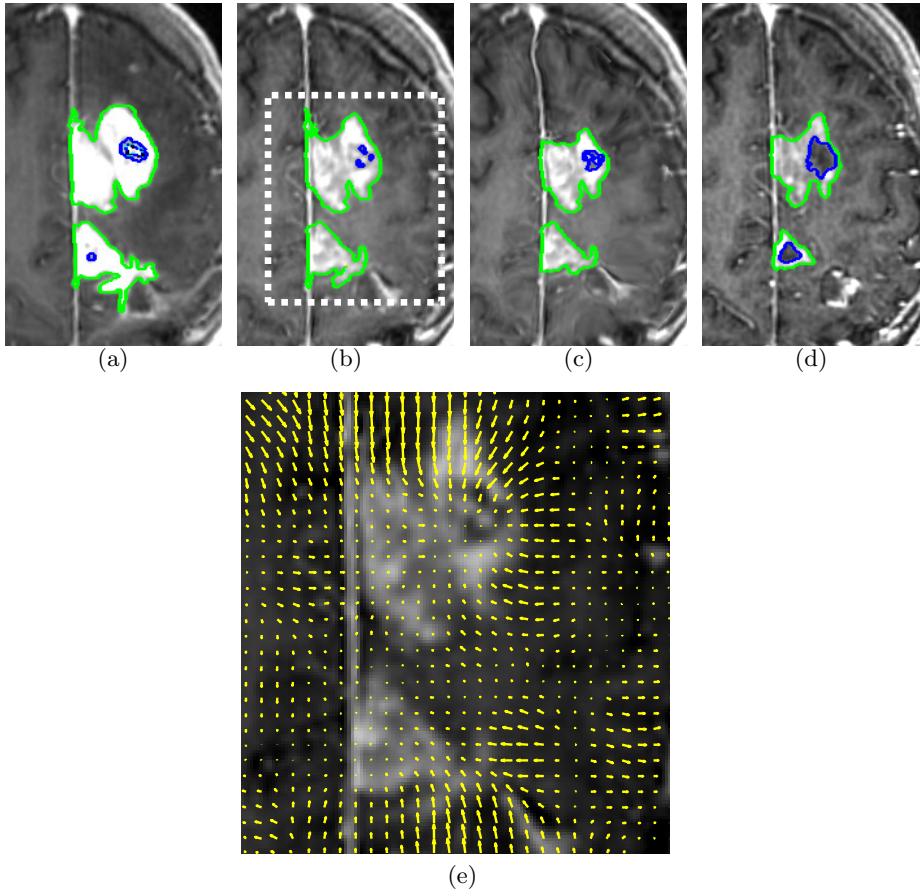
Fig. 2. Illustration of the method on a representative slice of T1-weighted pulse sequence of an GBM tumor and its internal components: (a) past scan; (b) current scan; (c) predicted scan; (d) actual acquired future scan – the components are: enhancing (green) and necrotic (blue); (e) estimated future transformation field projected on the scan plane and overlaid on the dashed area of the current scan – the yellow arrows indicate the local direction of the estimated future transformation field. Note how the algorithm successfully predicted the regression of the enhancing part of the tumor and the enlargement of the necrotic area.

(a)   (b)   (c)
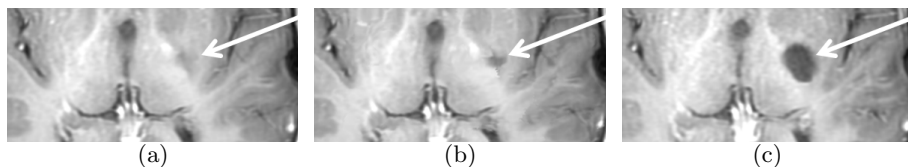
Fig. 3. A representative slice of T1-weighted pulse sequence from (a) current scan; (b) predicted scan, and; (c) actual acquired future scan. The arrow points to a new region of involvement of the tumor. Note that the new area of involvement is clearly noticeable in the predicted scan, although it is barely seen in the current scan.

Table 1. Volumetric overlap and mean surface distance of manual segmentation results: Current and predicted scans vs. actual future scan. The p-values were computed with the Wilcoxon signed rank test.

	Patient #	Volumetric overlap error [%]		Mean surface distance [mm]	
		Predicted vs. Actual future	Current vs. Actual future	Predicted vs. Actual future	Current vs. Actual future
OPG patients	1	32.9	32.8	1.21	1.15
	2	9.4	13.1	0.13	0.2
	3	8.3	23.3	0.22	0.75
	4	4	5.8	0.03	0.05
Average (OPG)		13.65	18.75	0.4	2.15
p-value (OPG)		0.25		0.375	
GBM patients	5	21.2	84.7	0.73	51.88
	6	25.1	65.1	0.18	1.47
	7	54.3	65.3	1.35	1.68
	8	36.3	38.5	1.038	1.29
Average (GBM)		34.23	63.4	0.83	14
p-value (GBM)		0.125		0.125	

We conclude from Table 1 that the VOE between the predicted segmented tumor vs. the actual future tumor is in many cases substantially lower that the one between the current and actual future ones. This indicates that when the method successfully estimated the future growing trend of the tumor, the predicted scan is more similar to the actual future scan than the current scan is.

A key new feature of our method is its ability to expand and emphasize new regions of tumor involvement during their early stage, in which they can be easily missed by the radiologist. Consider for example Fig. 3, which shows a representative slice from the current, predicted, and actual future scans. Note that while the infiltration of the cystic component into the slice can hardly be noticed in the current scan, this trend is emphasized in the predicted scan. This provides a clear early indication for this future involvement area of the tumor.

4   Conclusions

We have presented a new method for the estimation of patient's future MR scan based on his/her past scans based on an auto-regression model of the voxel's

optical flow over time. Since our method does not rely on a tumor-specific growth model, it can be used for a wide variety of tumor types. Experimental results on two types of brain tumors, OPG and GBM, show the successful prediction of the tumor growth, in terms of volumetric tumor overlap with the segmentation of the actual future scan.

A key advantage of our method is that it is fully automatic and that it does not require any user intervention for tumor localization. Furthermore, it predicts a whole scan rather than a specific tumor region of interest. Therefore, it has the potential to emphasize, in the predicted scan, tumor regions in their initial stage, that can hardly be noticed in the current scan.

Our method has several limitations. First, it cannot handle a surgical intervention performed between scans, as the prediction based on scans before the surgery most likely will hamper the prediction process. Second, the autoregressive assumption might not hold for all cases. In these cases, a specific tumor growth model has to be taken into account for a better prediction. Finally, the method would fail to predict substantial changes in tumor growth, which have no evidence in the past, or cases where the future trend of tumor growth is opposite to the one seen in the past.

The potential clinical significance of the brain tumor prediction is to provide the clinicians with a better picture of the tumor growth trends. The predicted scan can then be used to detect future involvement area of the tumor and to assist the decision-making process. Our future work will include the generalization of the method to include other MR sequences in the prediction process. In addition, we will examine the methods' performance on other types of tumors.

References

1. Hogea, C., Davatzikos, C., Biros, G.: Modeling Glioma Growth and Mass Effect in 3D MR Images of the Brain. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 642–650. Springer, Heidelberg (2007)
2. Clatz, O., Sermesant, M., Bondiau, P.Y., Delingette, H., Warfield, S., Malandain, G., Ayache, N.: Realistic simulation of the 3D growth of brain tumors in MR images coupling diffusion with biomechanical deformation. IEEE Transactions on Medical Imaging 24(10), 1334–1346 (2005)
3. Mardor, Y., Roth, Y., Ocherashvilli, A., Spiegelmann, R., Tichler, T., Daniels, D., Maier, S.E., Nissim, O., Ram, Z., Baram, J., Orenstein, A., Pfeffer, R.: Pretreatment prediction of brain tumors' response to radiation therapy using high b-value diffusion-weighted MRI. Neoplasia 6(2), 136–142 (2004)
4. Loncaster, J.A., Carrington, B.M., Sykes, J.R., Jones, A.P., Todd, S.M., Cooper, R., Buckley, D.L., Davidson, S.E., Logue, J.P., Hunter, R.D., West, C.M.: Prediction of radiotherapy outcome using dynamic contrast enhanced MRI of carcinoma of the cervix. International Journal of Radiation Oncology Biology Physics 54(3), 159–767 (2002)
5. Friston, K.J., Holmes, A.P., Ashburner, J.: Statistical parametric mapping, SPM (1999), http://www.fil.ion.ucl.ac.uk/spm/
6. Kroon, D.J., Slump, C.: MRI modalitiy transformation in demon registration. In: IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 963–966 (2009)

7. Thirion, J.: Image matching as a diffusion process: an analogy with maxwells demons. Medical Image Analysis 2(3), 243–260 (1998)
8. Morettin, P.A.: The Levinson algorithm and its applications in time series analysis. International Statistical Review / Revue Internationale de Statistique 52(1), 83–92 (1984)
9. Gerig, G., Jomier, M., Chakos, M.: Valmet: A New Validation Tool for Assessing and Improving 3D Object Segmentation. In: Niessen, W., Viergever, M. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 516–523. Springer, Heidelberg (2001)

Resting-State FMRI Single Subject Cortical Parcellation Based on Region Growing

Thomas Blumensath[1], Timothy E.J. Behrens[1,2], and Stephen M. Smith[1]

[1] Oxford Centre for Functional MRI of the Brain (FMRIB Centre),
University of Oxford, Oxford, UK
{tblumens,behrens,steve}@fmrib.ox.ac.uk
www.fmrib.ox.ac.uk
[2] Welcome Trust Centre for Neuroimaging, 12 Queen Square, London, UK

Abstract. We propose a new method to parcellate the cerebral cortex based on spatial dependancy in the fluctuations observed with functional Magnetic Resonance Imaging (fMRI) during rest. Our surface-based approach uses a region growing method. In contrast to previous methods, locally stable seed points are identified on the cortical surface and these are grown into a (relatively large 1000 to 5000) number of spatially contiguous regions on both hemispheres. Spatially constrained hierarchical clustering is then used to further combine these regions in a hierarchical tree. Using short-TR resting state fMRI data, this approach allows a *subject specific* parcellation of the cortex into anatomically plausible subregions, identified with high scan-to-scan reproducibility and with borders that delineate clear changes in functional connectivity.

1   Introduction

A subdivision of the human cerebral cortex into anatomically and functionally distinct regions has long been a fundamental goal in the study of the human brain. Measuring Blood Oxygen Level Dependant (BOLD) signal fluctuations during rest using functional MRI (fMRI) techniques allows studies of such functional subdivisions in the living human brain. We propose a novel automated procedure to derive such a partition (or parcellation) for an *individual subject* using resting state BOLD data. Not only is this single subject approach able to uncover fundamental organisational principles, but is also anticipated to better reveal subject specific variations in human functional brain anatomy. Furthermore, the robust specification of cortical parcellations is a fundamental first step in the definition and study of functional brain networks [1], which model statistical dependencies between the parcel's characteristic time-courses.

2   Methodology

A wide range of statistical tools has been used over the years to derive cortical parcellations from resting state fMRI data. Methods include those based

on ICA, spectral clustering, hierarchical clustering, k-means (or fuzzy k-means) clustering, mixture model based clustering and many more [2,3,4]. Region growing approaches for fMRI-based parcellation have previously been suggested in [5,6,7]. Our approach uses the following steps:

1. We use ICA based automatic de-noising of the data set.
2. We perform our analysis on the cortical surface.
3. We define a fairly large number of 'locally stable' seeds.
4. The seeds are then grown into spatially contiguous regions.
5. A cluster tree is build using spatially constrained hierarchical clustering.

2.1 Subjects and Data Acquisition

We used data from a previous study [8]. For each of the 5 subjects (ages 18-25, 3 males and 2 females) six 10-miniute fMRI datasets were acquired in a single session (eyes closed) using the accelerated protocol describe in [9], providing whole brain coverage at a TR of 0.8 seconds and with an isotropic spatial resolution of 3mm. Data was acquired on a 3 Tesla standard commercial scanner (Siemens Trio) equipped with 40 mT/m gradients with a slew rate of 200 mT/m/ms. In each scanning session, a single EPI reference image was acquired between each of the 10 minute blocks. A 1x1x1 mm^3 resolution structural image (T1-weighted) was acquired to aid registration. One of the subjects was scanned 3 times, on separate days, so the full data-set consisted of 42 10 minute sections. Of these, 36 were no-task resting-state paradigms (30 with eyes closed, 6 with eyes open). The remaining 6 scans (eyes open) used a silent backwards counting task. The imaging protocol used for human studies was approved by the institutional review board at the University of Minnesota. Each of the subjects provided informed written and verbal consent prior to participating in the research.

2.2 Preprocessing

FSL (FMRIB's Software Library) [10] was used for 1) head motion correction 2) full width 200s temporal high-pass filtering and 3) ICA de-noised with spatial-ICA using MELODIC [11]. 4) T1 images were segmented using FreeSurfer [12] and the functional data was projected onto the cortical surface using FreeSurfer's mri_vol2surf functionality.

2.3 Defining Locally Stable Seeds

BOLD signal to noise ratio (SNR) is not uniform across the cortical surface. However, there are currently no methods that would allow us to estimate this varying SNR reliably. Nevertheless, measures used to assess functional connectivity will indicate a decreased connectivity if spatially independent noise is present, so that brain areas with higher noise contamination will in general show lower connectivity. To partially address the SNR problem, we therefore propose an approach that is based on the estimation of cortical locations that can be assumed to be less noisy.

A further constraint we would like to impose on these locations is that they are placed well inside a functionally homogeneous brain region. We calculate a spatial map that encodes both of these properties. We here computed the local standard deviation in a 3mm radius, that is we subtract the ROI's mean time course from each of the ROI's time-courses and then calculate the standard deviation of the concatenated demeaned time-courses. Larger ROI radii gave similar results (though this produced smoother maps and thus fewer seed points). Seed points were found to be distributed approximately uniformly over the cortex.

To further reduce the variance of the estimated map (at the expense of spatial resolution), we post-process the map using a surface based Gaussian kernel smoother with a full width half max of 2.35mm ($\sigma = 1mm$). Seed voxels are then found as the local minima within the smoothed map.

2.4   Growing the Seeds

Seeds are grown into different clusters using region growing. Pearson correlation is calculated between the seed's time-course and the time-courses of its neighbouring vertices. Neighbouring vertices are then joined to a seed's cluster if their similarity measure is within 90% of the maximal similarity found within this iteration. The 90% threshold was chosen to provide a good compromise between speed (higher thresholds require more iterations) and performance (lower thresholds increase the possibility of suboptimal assignment). This step is repeated iteratively until all vertices are assigned. For each cluster, the neighbouring vertices not yet assigned to a cluster are selected and their similarity to the neighbouring cluster's seed time-courses is calculated and vertices are again assigned to clusters if their time-course similarity is within a fraction of the maximum such similarity.

2.5   Clustering the Seed Regions Using a Spatial Constraint

The number of seeds depends to some extent on the size of the neighbourhoods used to define the original stability map as well as on the amount of smoothing used. The parameters described above led to a smoothness that produced approximately 1000 seeds on each cortical hemisphere, though we also tried less smooth maps, which produced roughly 3000 seeds per hemisphere (final clustering results were marginally worse when using the larger number of seeds).

Whilst there is no 'correct' resolution, the number of seeds produced by the above approach is still much larger than that used typically in functional connectivity studies or expected from a cytoarchitectonic atlas. We thus devised a spatially constrained hierarchical clustering approach to further combine regions hierarchically. The spatial constraint ensured that at each resolution in the cluster tree, the parcellation consists of spatially connected regions.

We calculated Pearson's correlation between all of the regions time-courses and also defined a neighbourhood structure between clusters. Clusters are then joined iteratively by 1) joining two neighbouring clusters whose similarity was maximal among all neighbouring clusters; 2) updating the similarity profile for

the joint clusters and their neighbours using a linkage rule. We use the linkage function proposed by Ward [13], even though Ward's method assumes Euclidean distances instead of correlations. Importantly, this procedure produces an entire cluster tree and not just a single parcellation. The tree can then be cut at a desired resolution, depending on the application.

3   Results

Evaluating parcellation methods is difficult, as no ground truth is available. We thus here look at several necessary conditions a good cortical parcellation should satisfy. The first requirement is that the method should produce similar parcellations when run on different data sets derived from the same subject. We will use this as our main measure of the quality of the parcellation, though it is important to realise that this measure is not sufficient to characterise neuro-anatomically meaningful parcellations. For example a parcellation approach might be biased in terms of the size or shape of parcels it returns and many of these biases would also increase scan-to-scan reproducibility artificially.

3.1   Scan to Scan Reproducibility

In our data set we have two sets of 6 10-minute scans acquired from the same subject under identical conditions. We used this data-set to generate two parcellations. To reduce potential variability in the results due to scanning day or order effects, we used the first, third and fifth scan from session one together with the second, fourth and sixth scan from session two to generate the first parcellation. The second parcellation was derived from the remaining six scans.

Similarity between two parcellations was measured using the dice coefficient. Instead of measuring the dice similarity between any two pairs of clusters in two parcellations, we first matched clusters. This was done by matching the two clusters that have the largest dice similarity. This process was repeated (with already matched clusters removed from further comparison). Dice similarities between matched pairs were then averaged to get a scalar measure of parcellation similarity.

The left panel in Figure 1 plots the dice similarity (y-axis) for parcellations obtained by cutting the hierarchical cluster tree at different resolutions (x-axis). We also show the results we obtained with other approaches in Figure 1. See the caption for details and references.

The only approach which we found to performed similarly to our method was a locally constrained normalised cuts (NCUTS) spectral clustering method where the similarity matrix was constrained to spatial neighbourhoods. However, looking at the parcellations found with this approach (shown in Figure 2(b)), it appears that this method produced clusters that are less anatomically convincing than those derived with our approach (shown in Figure 2(a)). In particular, NCUTS completely failed to identify the sensory-motor cortex, which is parcellated into body and face regions with the region growing method.
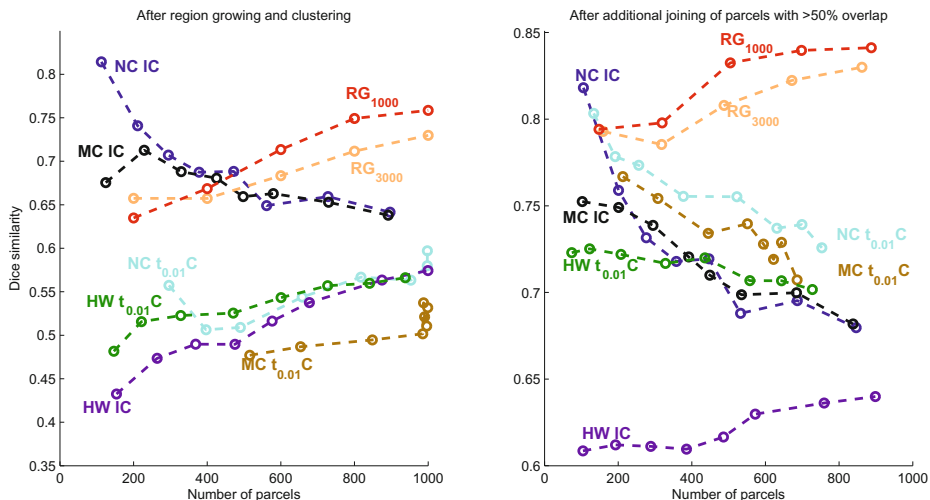
Fig. 1. Dice similarity between matched parcellations vs. parcellation size calculated for parcellations from different datasets of the same subject before (left) and after (right) joining split clusters. Region growing approach with ~1000 (RG_{1000}, red) and ~3000 (RG_{3000}, orange) seeds outperforms all other tested approaches over a range of parcellation resolutions. Also shown is a small subset of the results obtained with other methods (including the next best performing methods NCUTS with a local correlation similarity measure (NC LC, blue) [4] a spectral clustering approach to optimise network modularity (MC LC, black) [14] and a hierarchical clustering approach using Ward's linkage rule (HW LC) [13], both again with a locally restricted correlation similarity measure. For comparison, the results obtained by the same approaches with a different similarity matrix (a correlation matrix in which values were thresholded so that only 1% of the entries in each column/row where non-zero [2]) are also shown ({NC, MC, HW} $t_{0.01}$C).)

The stark difference between the results obtained with our approach and those found with NCUTS seems to be mainly due to NCUTS being strongly biased towards a parcellation with clusters of similar size (compare the histogram inlays in figure 2). Note that the effect of this bias in NCUTS reduces if the number of parcels increases (hence the dice measure decreases for NCUTS in figure 1 with increasing resolution, whilst dice incases for all other methods).

Due to noise effects, a region is sometimes split into two parcels in one parcellation, whilst in another parcellation the regions have been joined. To account for this effect, when comparing parcellations, we produced matched parcellations in which any two clusters in one parcellation that each overlapped more than 50% with a cluster in the second parcellation, were joined iteratively. Dice similarity after this joining operation is shown in the right panel of Figure 1. Also note that the parcellations shown in Figure 2 are those found after this joining operation and that colours were matched to ease comparison.

(a) Our region growing approach

(b) NCUTS with locally constrained correlation [4]

Fig. 2. Same subject parcellation with joint clusters from data-set 1 (left) and dataset 2 (right). Histogram inlays show the distribution of parcel sizes. Parcel colours have been matched to ease comparison.

3.2 Borders Represent Changes in the Connectivity Profile

Finally, to show that the borders of the derived parcellations indicate changes in functional connectivity, we drew a path[1] along the cortical surface (shown on the left in Figure 3). Using a *different* data-set (from the same subject) to that used to derive the parcellation itself. we calculated correlations between the time-series of each vertex along the path (the correlation matrix (thresholded at ± 0.4) is the lower left part of the matrix on the right of Figure 3) and also calculated a connectivity profile feature vector for each vertex along the path (this feature vector was the r to z transformed correlation between the vertex time-series and the time-series of 1000 randomly chosen but fixed target vertices on the cortical surface). The correlation between these feature vectors is shown as the upper triangular part of the matrix. Parcellation borders (black lines) align with significant changes in the connectivity profile.

---

[1] The path was chosen so that it roughly crosses parcellation borders at a right angle.
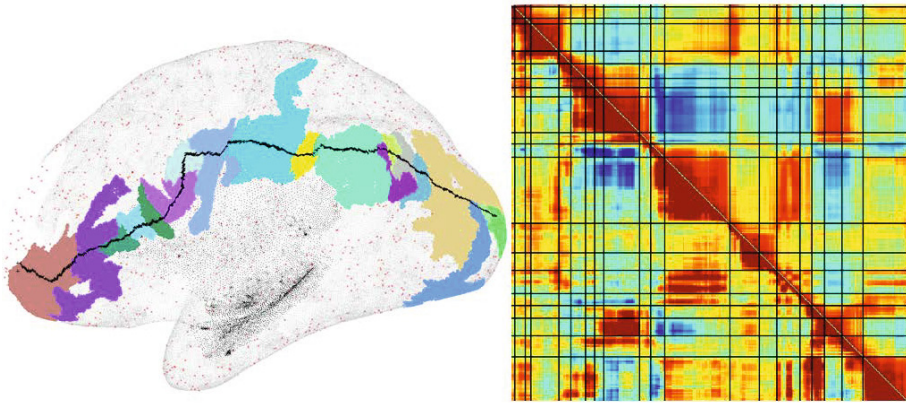
Fig. 3. Path drawn on the cortical surface together with the parcels the path crosses (left) and the correlation (lower left) and connectivity (upper right) profile along the path with parcel borders shown as black lines (right).

4   Conclusions

A subject specific cortical parcellation based on resting state functional MRI data has many applications in the study of functional neuro-anatomy and functional brain connectivity. We have proposed an approach that shows high scan-to-scan reproducibility and was able to derive parcellation borders that clearly follow changes in the functional connectivity profile. Comparison to other approaches demonstrates the superiority of the proposed approach. Whilst single subject based methods are desirable, they also remain limited in the quality achievable with a single subject's data set and even using 60 minutes of high quality data as done here, parcellations do not yet reach the resolution and reliability desirable for detailed anatomical study. To overcome this, we are currently working on approaches that stabilise single subject parcellations using multi-subject data-sets in a more robust way than is achievable currently with registration based on gross anatomical features.

References

1. Smith, S.M., Miller, K.L., Salimi-Khorshidi, G., Webster, M., Beckmann, C.F., Nichols, T.E., Ramsey, J.D., Woolrich, M.W.: Network modelling methods for FMRI. NeuroImage 54(2), 875–891 (2011)

2. Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., Petersen, S.E.: Functional network organization of the human brain. Neuron 72(4), 665–678 (2011)
3. Mumford, J.A., Horvath, S., Oldham, M.C., Langfelder, P., Geschwind, D.H., Poldrack, R.A.: Detecting network modules in fMRI time series: a weighted network analysis approach. Neuroimage 52(4), 1465–1476 (2010)
4. Craddock, R.C., James, G.A., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. Hum. Brain Mapp. (July 2011)
5. Lu, Y., Jiang, T., Zang, Y.: Region growing method for the analysis of functional mri data. NeuroImage 20(1), 455–465 (2003)
6. Bellec, P., Perlbarg, V., Jbabdi, S., Plgrini-Issac, M., Anton, J.L., Doyon, J., Benali, H.: Identification of large-scale networks in the brain using fMRI. NeuroImage 29(4), 1231–1243 (2006)
7. Heller, R., Stanley, D., Yekutieli, D., Rubin, N., Benjamini, Y.: Cluster-based analysis of fmri data. NeuroImage 33(2), 599–608 (2006)
8. Smith, S.M., Miller, K.L., Moeller, S., Xu, J., Auerbach, E.J., Woolrich, M.W., Beckmann, C.F., Jenkinson, M., Andersson, J., Glasser, M.F., Van Essen, D.C., Feinberg, D.A., Yacoub, E.S., Ugurbil, K.: Temporally-independent functional modes of spontaneous brain activity. Proceedings of the National Academy of Sciences 109(8), 3131–3136 (2012)
9. Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E.: Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. PLoS ONE 5(12) (December 2010)
10. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., Luca, M.D., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23(S1), 208–219 (2004)
11. Beckmann, C., Smith, S.: Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE Transactions on Medical Imaging 23(2), 137–152 (2004)
12. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. NeuroImage 9(2), 179–194 (1999)
13. Ward, J.H.J.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301), 236–244 (1963)
14. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 036104 (2006)

A Framework for Quantifying Node-Level Community Structure Group Differences in Brain Connectivity Networks

Johnson J. GadElkarim[1,2], Dan Schonfeld[1], Olusola Ajilore[2], Liang Zhan[3], Aifeng F. Zhang[2], Jamie D. Feusner[4], Paul M. Thompson[3], Tony J. Simon[5], Anand Kumar[2], and Alex D. Leow[2,6,*]

[1] Electrical and Computer Engineering department, University of Illinois at Chicago
[2] Department of Psychiatry, University of Illinois at Chicago
[3] Laboratory of Neuro Imaging, UCLA School of Medicine
[4] Department of Psychiatry and Behavioral Sciences, UCLA
[5] Department of Psychiatry and Behavioral Sciences, UC Davis
[6] Community Psychiatry, Sacramento, CA
jgadel2@uic.edu, alexfeuillet@gmail.com

Abstract. We propose a framework for quantifying node-level community structures between groups using anatomical brain networks derived from DTI-tractography. To construct communities, we computed hierarchical binary trees by maximizing two metrics: the well-known modularity metric (Q), and a novel metric that measures the difference between inter-community and intra-community path lengths. Changes in community structures on the nodal level were assessed between generated trees and a statistical framework was developed to detect local differences between two groups of community structures. We applied this framework to a sample of 42 subjects with major depression and 47 healthy controls. Results showed that several nodes (including the bilateral precuneus, which have been linked to self-awareness) within the default mode network exhibited significant differences between groups. These findings are consistent with those reported in previous literature, suggesting a higher degree of ruminative self-reflections in depression.

1 Introduction

Modern imaging techniques have allowed us to construct anatomical (structural) and functional networks of the human brain and a new term - the human "connectome" [1] has been recently coined to describe their mathematical properties. A brain connectivity network consists of nodes which represent gray matter regions and edges connecting nodes which represent white matter fibers. Many tools and metrics have been adapted from the field of graph-theory to help characterize and analyse the human connectome [1-2]. It has been shown that brain networks are sparse (networks having the number of nodes in the order of the number of edges)

---

[3], thus allowing the investigation of their community structure properties in which nodes may be grouped into different communities [4, 5] (in our context, "*community structure*" is a property for sparse networks, which allows us to extract the topological organization of a network and to partition it into a set of non-overlapping *communities*, sometimes also referred to as *modules* or *clusters*).

Studying the structural organization of the brain communities has lately shed light on alterations related to aging [6, 7] as well as neuropsychiatric disorders such as schizophrenia [8-12]. In this paper, we proposed a novel metric, with which we extracted community structures in the form of top-down binary trees (also known as dendrograms). This is followed by the computation of a node "consistency metric", designed to quantify differences between trees at each node. To our best knowledge, we are the first group to propose a recipe for quantifying node-level community differences between two groups of brain networks. Lastly, statistical analyses were conducted to detect significant community alterations on the nodal level for group studies. We tested the proposed method and compared the binary trees generated from the structural brain networks of depressed versus normal control subjects. Significant changes were found on the nodal level in the community structure in depression relative to controls.

2   Methods

2.1   Image Acquisition

We scanned 47 healthy subjects (20 male/27 female; age: 59.74 ± 14.8)) and 42 subjects with major depression as defined by DSM-IV (18 male/24 female; age: 57.64 ± 13.4). A Philips 3.0 T Achieva scanner supplied with 8-channel SENSE head-coil located at the University of Illinois Medical Center Advanced Imaging Center was used to acquire the brain MRI data. High resolution T1-weighted images were acquired with MPRAGE sequence (FOV=240 mm; TR/TE=8.4/3.9 ms; flip angle 8°; voxel size 1.1x1.1x1.1mm). Diffusion weighted (DW) images were acquired using SS-SE-EPI sequences (FOV = 240 mm; resolution 0.83X0.83 mm; TR/TE=6994/71 ms; flip angle = 90°, 32 gradient directions, b= 700 s/mm^2 and one minimally DW scan: b0 image). Parallel imaging was also used with a SENSE factor of 2.5 to reduce scan time to $\sim$ 4 min.

2.2   Brain Network Construction

Structural brain networks were generated using a pipeline which integrates multiple image analysis techniques. First, DW images were eddy current corrected using the automatic image registration (AIR) tool embedded in DtiStudio software (http://www.mristudio.org) by registering all DW images to their corresponding b0 images with 12-parameter affine transformations. This was followed by the computation of diffusion tensors and deterministic tractography using the DtiStudio program. T1-weighted images were used to generate label maps using the Freesurfer software (http://surfer.nmr.mgh.harvard.edu). Brain networks formed by the 68 cortical regions were generated using an in-house program in Matlab by counting the number of fibers connecting each pair of nodes.

2.3   Hierarchical Trees Creation

Measuring community structure has been a scientifically important task in network science. To this end, we note that nodes belonging to the same community should exhibit stronger connections than those in different communities. A well-known metric for such task is the Q modularity proposed in [13] and is mathematically defined as:

$$Q\left(G\right) = \frac{1}{2m} \sum_{i \neq j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta\left(i,j\right) \tag{1}$$

where Q is a function of a graph G, m is the total number of edges, $A_{ij} = 1$ if an edge links nodes i and j and 0 otherwise, $\delta(i, j) = 1$ if nodes i and j are in the same community and 0 otherwise, and k_i is the node i's degree (its number of edges). Extracting the community structure of a network is obtained by finding the set of non-overlapping modules that maximizes Q. Here, we further propose an alternative maximization problem with respect to a novel metric. This new metric is defined as the difference between the mean inter- and the mean intra-modular path lengths ($inter_{PL}$/$intra_{PL}$). For two communities C_i and C_j the $inter_{PL}$ and $intra_{PL}$ are defined as:

$$inter_{PL}^{C_i \leftrightarrow C_j} = \frac{\sum_{n \in C_i, m \in C_j} d_{nm}}{N_i N_j}, intra_{PL}^{C_i} = \frac{\sum_{n,m \in C_i; n > m} d_{nm}}{\left(N_i^2 - N_i\right)/2} \tag{2}$$

where N_i denotes the number of nodes in a community C_i, d_{nm} denotes the shortest path length connecting nodes n and m. In the case of two modules, for example, the alternative metric Ψ^{PL} is defined as:

$$\Psi^{PL} = inter_{PL}^{C_1 \leftrightarrow C_2} - \frac{1}{2} \left[intra_{PL}^{C_1} + intra_{PL}^{C_2} \right] \tag{3}$$

Here, notice that $intra_{PL}^{C_i}$ can be considered the weighted characteristic path length (a measure of global network integration) defined on the sub-network formed by C_i. Thus, maximizing Ψ^{PL} is equivalent to searching for a set of communities that exhibit stronger sub-network integration. Since path length is based on the number of fibers, Ψ^{PL} may provide a more anatomical basis for defining community structure.

There exist two main types of hierarchical clustering techniques: the agglomerative and the divisive. The agglomerative method starts bottom-up. At the first level, every node is a community by itself. In each subsequent step (level), two communities that are considered closest (with respect to a similarity measure) are merged into one, until all nodes are in one community. The divisive method follows the opposite or a top-down approach. It starts with all nodes belonging to one single community, and at each step each community is further split into several sub-communities (unlike the agglomerative method, the divisive clustering usually uses a dis-similarity measure). It should be noted that one may choose to stop the clustering process at a pre-specified level (stopping criteria can be a given number of communities or the optimization of a quality function

[14]). Here, to maximize Q or Ψ^{PL}, we used a top-down hierarchical clustering approach, as it has been suggested that agglomerative methods may fail to find correct communities in networks where the community structures are known a priori [13]. At the first level, brain regions were randomly assigned to one of two communities, and optimal assignment was determined by maximizing Q or Ψ^{PL} using the well-known simulated-annealing algorithm [15]. The process is repeated at each level until a 4-level dendrogram is reached (a total of 16 communities). One million random permutations were used at each optimization stage.

2.4   Community Structures Difference Assessment

In order to assess local differences between communities in different groups, we first construct the mean binary tree for the control group (by maximizing Q or Ψ^{PL}, with respect to the mean connectivity matrix for this group). All individual subjects' trees will then be compared to the mean control tree. To this end, a local metric is needed to quantify how two trees differ. One candidate metric is via the consistency metric assessment method introduced in [16], which yields a consistency vector V of length equal to the number of nodes in the network (68 in our case). To compute V, we need to first construct a LxM similarity matrix X_R that compares any module in a test tree (i.e., the individual subject's tree in our framework) to any module in a reference tree (the mean tree for the control group). Here L and M are the number of communities in the test and reference trees; in our case L = M = 16. Mathematically, for any two communities C_p and C_q belonging to the test tree and the reference tree respectively, the $(p, q)^{th}$ entry $X_R(p, q)$ in X_R is calculated using the following equation:

$$X_R(p, q) = \frac{N_{pq}^2}{N_p N_q}; p = 1, ..., L; q = 1, ..., M \tag{4}$$

where N_{pq} denotes the number of common nodes between the two communities C_p and C_q, and N_p and N_q the number of nodes in community C_p and C_q. Values in X_R range from 0 to 1, with 0 indicating no overlap between the two communities and 1 identical communities (numerically, $X_R(p, q)$ thus represents the similarity between community C_p in the test tree and community C_q in the reference tree). With the computation of X_R, the consistency vector V now can be constructed as follows. First, we identify the maximum entry in each column of the X_R matrix. For the qth column corresponding to community C_q in the reference tree, if the maximum exists in the pth row corresponding to community C_p in the test tree, we assign this value to the kth element of V, for all $k \in \{C_p \cap C_q\}$. Although V provides a node-level measurement of how "consistent" the test tree's community structure is with respect to that of the reference tree, it should be noted that with this definition, V may yield zero values. We thus proposed a modified non-zero node-level consistency measure (V^m) for each node k as follows:

$$V^m(k) = \frac{(N_c)^2}{N_p N_q}, k = 1, ..., 68 \tag{5}$$

where N_c denotes the number of common nodes between the two communities C_p and C_q, that contain this node k in the test and reference trees respectively.

2.5   Statistical Analysis

To statistically test group differences in community structures at the nodal level, all individual subjects' trees (both depressed and control) are compared to the mean control reference tree, thus yielding 47 V or V^m vectors in the healthy group, and 42 in the depressed group. Group differences on local community structures can now be assessed using node-wise 2-sample Mann-Whitney-Wilcoxon (MWW) tests for V (due to the existence of zero values) or node-wise 2-sample T-tests for V^m followed by correcting for multiple comparisons. Alternatively, a more powerful test can be constructed via a multivariate distribution for each community in the 4th level binary tree of the mean normal group (16 total modules), by concatenating the V or V^m vectors of all nodes in this community, on which 2-sample Hotelling's T-squared tests can be conducted.

3   Results

3.1   Comparison of Modularity Measures

We first generated and compared the mean community structures of the normal controls in this study, using both Q and Ψ^{PL} (figure 1). Note that in the first level of the binary tree (partitioning the brain into two modules), the constructed mean tree using the Q function exhibits an exact left/right separation while the tree generated using Ψ^{PL} shows an anterior/posterior partition of the human brain (the left-right separation represents the second level partitioning using the proposed metric; the mean community structure of the studied depressed subjects is shown in right panel of figure 1). The second level partitioning using Ψ^{PL} also generated more unified communities compared to the Q function.

3.2   Group Difference

To detect community structure abnormalities in depression, we constructed the Hotelling T-squared statistics as described in the method section for all 16 communities in the reference tree. Results showed a significantly lower consistency (p = 0.0079) in depression versus control for the community containing the right precuneus, superior parietal gyrus, inferior parietal gyrus, inferior temporal gyrus, and isthmus cingulate (figures 2 and 3). On a node level, conducting 2-sample T-tests or the MWW-test on V^m or V confirmed such findings (table 1). The right precuneus and superior parietal gyrus exhibited the lowest node-level consistency (p < 0.01; uncorrected). Comparing the mean community structures between two groups, we note that the right precuenus and right superior parietal gyrus, while belonging to the same community in the healthy controls, are assigned to different communities in the depressed group. Additionally, in depression the bilateral precuneus are assigned to one community, suggesting a stronger structural integration between them (figure 3).
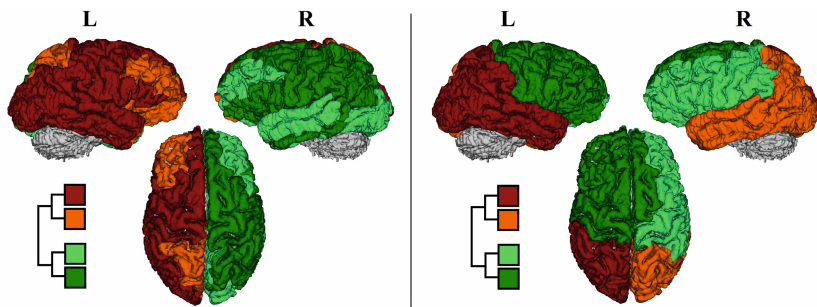
Fig. 1. First two levels of modular or community structure via the Q modularity (left) and Ψ^{PL} (right). Notice that at the first level, the constructed mean tree by maximizing Q exhibits a left/right parcellation while the constructed mean tree using Ψ^{PL} exhibits an anterior/posterior partitioning along the frontal-parietal junction.
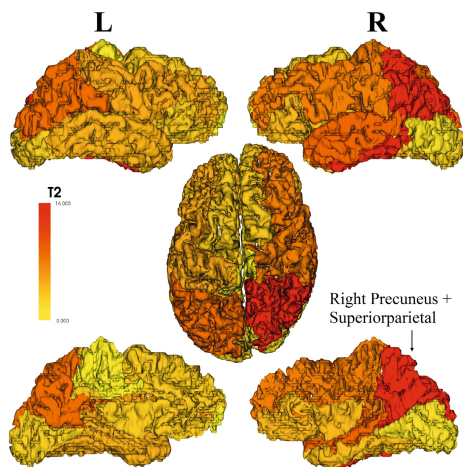
Fig. 2. Hotelling's T-Squared statistics are overlaid on the 16 communities of the reference tree, showing regional group differences in the community structure between groups. The most significant differences are in the right precuenus and superior parietal gyrus. This figure is showing a left/right lateral view (top), an axial view (middle) and a left/right mid-line view (at the bottom). The community containing the right precuneus and the right superior parietal gyrus is highlighted in red.
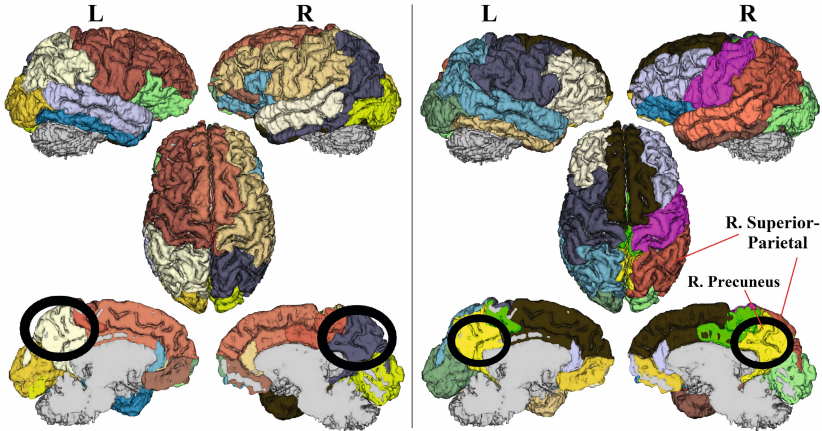
Fig. 3. 4-level binary trees showing mean community structures via the proposed metric Ψ^{PL} for the healthy controls (left) and the depressed subjects (right). Notice that the right precuneus and right superior parietal gyrus belong to the same community in the healthy controls but not in the depressed group. Additionally, in the depressed group the right and left precuneus are assigned to one single community (which also comprises bilateral isthmus cingulate).

Table 1. List of nodes which showed significantly lower node-level consistency in depression relative to healthy controls at $p < 0.01$ (uncorrected). The table shows mean and standard deviation values using both MWW (for V) and 2-sample T-tests (for V^m), and their corresponding p values.

		Healthy	Depressed	p-value
right-precuneus	MWW-test	0.2373 ± 0.1815	0.1281 ± 0.1311	0.0044
	T-test	0.2645 ± 0.1501	0.1784 ± 0.0908	0.0019
right-superiorparietal	MWW-test	0.2088 ± 0.1773	0.1007 ± 0.1376	0.0032
	T-test	0.2374 ± 0.1475	0.1524 ± 0.1106	0.0033

4 Discussion and Conclusion

In this paper we presented a new approach to construct hierarchical community structures from weighted structural brain networks. Our proposed metric took into account the path lengths within and between communities, and we used simulated annealing to construct top-down four-level binary trees. Moreover, we are the first to develop a statistical framework that allows for the quantification of node-level community structural differences between two groups of brain networks, via the consistency vector metric. This framework was validated on a sample of depressed patients. Results revealed local community differences in parts of the default mode network (especially the right precuneus), which has been previously linked to self-awareness and shown to exhibit abnormalities in depression [17, 18].

References

1. Sporns, O., Tononi, Ktter, R.: The human connectome: A structural description of the human brain. PLoS Computational Biology 1, 245–251 (2005)
2. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: Uses and interpretations. J. of NeuroImage 52, 1059–1069 (2010)
3. Bassett, B.S.: Small World Brain Networks. The Neuroscientist 12(6), 512–523 (2006)
4. Meunier, D., Lambiotte, R., Fornito, A., Ersche, K.D., Bullmore, E.: Hierarchical modularity in human brain functional networks. Front. Neuroinformatics 3, 37 (2009b)
5. Telesford, Q., Simpson, S., Burdette, J., Hayasaka, S., Laurienti, P.: The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain. Brain Connectivity 1(4), 295–308 (2011)
6. Meunier, D., Achard, S., Morcom, A., Bullmore, E.: Age-related changes in modular organi-zation of human brain functional networks. J. of Neuroimage 44, 715–723 (2009a)
7. Fair, D.A., Cohen, A.L., Power, J.D., Dosenbach, N.U., Church, J.A., Miezin, F.M., Schlaggar, B.L., Petersen, S.: Functional brain networks develop from a "local to distributed" organization. PLoS Comput. Biol. 5(5), e1000381 (2009)
8. Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., Jiang, T.: Disrupted small-world networks in schizophrenia. Brain 131, 945–961 (2008)
9. Bassett, D.S., Bullmore, E., Verchinski, B.A., Mattay, V.S., Weinberger, D.R., Meyer-Lindenberg, A.: Hierarchical organization of human cortical networks in health and Schiz-ophrenia. J. of Neuroscience 28(37), 9239–9248 (2008)
10. Lynall, M.E., Bassett, D.S., Kerwin, R., McKenna, P.J., Kitzbichler, M., Muller, U., Bullmore, E.: Functional connectivity and brain networks in schizophrenia. J. of Neuroscience 30(28), 9477–9487 (2010)
11. Van Den Heuvel, M.P., Mandl, R.C.W., Stam, C.J., Kahn, R.S., Hulshoff Pol, H.E.: Aberrant frontal and temporal complex network structure in schizophrenia: A graph theoretical analysis. J. of Neuroscience 30(47), 15915–15926 (2010)
12. Alexander-Bloch, A., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., Bullmore, E.: The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. J. of NeuroImage 59(4), 3889–3900 (2012)
13. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113 (2004)
14. Fortunato, S.: Community detection in graphs. Physical Reports 486(3-5), 75–175 (2010)
15. Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: Optimization by Simulated Annealing. Science 220(4598), 671–680 (1983)
16. Steen, M., Hayasaka, S., Joyce, K., Laurienti, P.: Assessing the consistency of community structure in complex networks. Physical Review. E 84, 016111 (2011)
17. Zhu, X., Wang, X., Xiao, J., Liao, J., Zhong, M., Wang, W., Yao, S.: Evidence of a Disso-ciation Pattern in Resting-State Default Mode Network Connectivity in First-Episode, Treatment-Naive Major Depression Patients. Biol. Psychiatry (2011) (Epub ahead of print)
18. Lou, H.C., Luber, B., Crupain, M., Keenan, J.P., Nowak, M., Kjaer, T.W., Sackeim, H.A., Lisanby, S.H.: Parietal cortex and representation of the mental Self. Proc. of the National Academy of Sciences of the United States of America 101 (2004)

A Feature-Based Developmental Model of the Infant Brain in Structural MRI

Matthew Toews[1], William M. Wells III[1], and Lilla Zöllei[2]

[1] Brigham and Women's Hospital, Harvard Medical School
{mt,sw}@bwh.harvard.edu
[2] Massachussetts General Hospital, Harvard Medical School
lzollei@nmr.mgh.harvard.edu

Abstract. In this paper, anatomical development is modeled as a collection of distinctive image patterns localized in space and time. A Bayesian posterior probability is defined over a random variable of subject age, conditioned on data in the form of scale-invariant image features. The model is automatically learned from a large set of images exhibiting significant variation, used to discover anatomical structure related to age and development, and fit to new images to predict age. The model is applied to a set of 230 infant structural MRIs of 92 subjects acquired at multiple sites over an age range of 8-590 days. Experiments demonstrate that the model can be used to identify age-related anatomical structure, and to predict the age of new subjects with an average error of 72 days.

1 Introduction

The human brain undergoes dramatic developmental changes in the first years of life. Structural MR imaging offers the potential to model these changes over time and across the human population, in order to understand normal growth patterns and assess potential disorders [1], for example neurodevelopmental disorders relating to pre-term birth [2,3]. Anatomical development is closely tied to chronological age, and computational tasks of interest include automatically learning links between anatomical image structure and age, and predicting the age or developmental stage of new subjects.

A number of methods are used to analyze aging in structural MRI of the adult brain, e.g. group analysis via voxel-based morphometry (VBM) [4], growth patterns of dilation and contraction [5], discriminative classifiers [6,7,8]. These methods generally require accurate intensity-based segmentation and registration, tasks which remain research challenges in the context of the infant brain [9,1] due to pronounced intensity changes over the course of development, e.g. the contrast inversion of white/grey matter during myelination [10]. Infant temporal atlases are thus often treated as templates constructed over relatively narrow age ranges via age-specific registration and segmentation methods [11,12,9], and quantitative analysis has been largely limited to measures of growth, e.g. tissue volume changes [2,3,13]. Modeling dynamic image measurements across the infant developmental age range remains a challenge.

This paper proposes modeling development as a collection of distinctive, conditionally independent image features, localized in space and time. Anatomical structure is modeled as persisting over limited spatial and temporal windows, and as potentially only occurring in subsets of subjects. This provides a natural means of describing phenomena such as the disappearance/emergence of structure, spatially varying developmental rates, natural inter-subject variability and pathology. The model is based on scale-invariant features [14,15], distinctive image patterns that can be robustly extracted in the presence of global image deformations including scale changes, and that can therefore offer an information source complementary to typical growth measures such as volume change. Our model builds on the approach of [16], where local image features are used to construct a binary classifier for Alzheimer's disease in T1 MRI data. Here, a posterior probability is defined over a continuous variable of age, conditioned on feature data. Anatomical structure is modeled as a latent variable conditioned on age, and hypothesis testing is used to identify features most informative regarding age. The model is trained from combined cross-sectional and longitudinal data, in order to identify age-informative anatomical structure, and age is predicted via maximum a-posteriori estimation.

Experiments demonstrate several important advancements on the state-of-the-art. The model can be automatically trained from a large set of infant T1-weighted MRI data, acquired at multiple sites and scanners. Spatially-localized, age-related anatomical patterns are discovered across the infant age range, including white matter myelination. Cross-validation trials predict subject age with an average error of 72 days, which to our knowledge is the first published result for automatic infant age prediction from structural MRI data.

2   Feature-Based Developmental Model

The proposed model hypothesizes the existence of anatomical features that can be localized in time and space, and used to represent development in structural MRI data. Scale-space theory provides a framework for identifying both the location and spatial extent of such features [14], and forms the basis for invariant feature detection methods widely adopted in the computer vision community [17,15]. So-called scale-invariant features are spherical image regions characterizing the 3D location x and scale σ of distinctive image patterns. They can be automatically extracted from images via Gaussian derivative operators, for instance as extrema in the difference-of-Gaussian (DoG) scale space [15]:

$$\{x_i, \sigma_i\} = \text{local} \underset{x,\sigma}{\arg\max} \left\{ |G(\sigma) * I(x) - G(\kappa\sigma) * I(x)| \right\}, \tag{1}$$

where $I(x)$ is the image, $G(\sigma)$ is a Gaussian kernel of variance σ^2, and κ is a constant defining the multiplicative sampling rate in scale. The DoG is a computationally efficient approximation to the Laplacian-of-Gaussian operator, which is effective in identifying natural blob-like image structures. Once identified, image content within each region (x_i, σ_i) is cropped and rescaled to a fixed-size region, then encoded as an appearance descriptor for computing feature-to-feature

correspondence. The gradient orientation histogram (GoH) descriptor has been shown to be among the most effective for image-to-image matching [17], here a 3D variant with 8 orientation and 8 spatial bins is computed from $11 \times 11 \times 11$ normalized voxel regions. Due to spatial normalization, the scale-invariant feature representation is independent of isotropic image scaling, and is thus particularly useful in characterizing local anatomy independently from global growth or volume change.

2.1   Probabilistic Model

Let $\bar{I} = \{I_j\}$ represent a set of local features extracted in a subject image I, and let A represent a random variable of age. The posterior probability of A conditioned on data $\bar{I}$ can be expressed using Bayes rule as:

$$p(A|\bar{I}) = p(\bar{I}|A)p(A)/p(\bar{I}), \tag{2}$$

where $p(\bar{I}|A)$ is the likelihood of A associated with data $\bar{I}$, $p(A)$ is the posterior probability of A, and $p(\bar{I})$ is a constant as data $\bar{I}$ are fixed. Under the assumption of conditionally independent local feature measurements, the likelihood can be expressed as:

$$p(\bar{I}|A) = \prod_j p(I_j|A), \tag{3}$$

where $p(I_j|A)$ is the likelihood associated with an individual observed feature I_j and age A.

An image set is naturally described in terms of distinctive local structure shared across images, for instance the human brain can be described by corpus callosum, ventricles, etc. The model here adopts a description based on a code book of distinctive image features. Let $F = \{f_i\}$ represent discrete random variable, where event f_i indicates a distinct mode of feature appearance and geometry. Applying marginalization and Bayes rule, the likelihood in Equation (3) is expressed as:

$$p(I_j|A) = \sum_{i \in F} p(I_j, f_i|A) = \sum_{i \in F} p(I_j|f_i, A)p(f_i|A). \tag{4}$$

In equation (4), $p(I_j|f_i, A)$ represents the likelihood of feature f_i and age A associated with I_j, which can generally be taken to be a unimodal Gaussian density over feature parameters of location, log scale and appearance descriptor elements. Factor $p(f_i|A)$ is the conditional probability of f_i given age A, and is modeled as a multinomial distribution.

2.2   Learning, Analysis, Fitting

The goal of model learning is to generate a code book $F = \{f_i\}$ of model features characteristic of a training image set, and to estimate associated age-related factors $p(I_j|f_i, A)$ and $p(f_i|A)$. Prior to learning, training images are

normalized to a common reference space via a global coordinate transformation to an atlas template, e.g. a similarity or affine transform. Scale-invariant features are then extracted from each image, and a code book $F = \{f_i\}$ of model features representative of the image data is generated. A number of approaches could be used for this purpose, here we adopt a robust clustering strategy [16] similar to the mean-shift algorithm [18], which identifies clusters of features that are similar in terms of geometry and appearance across subjects. Each cluster f_i is characterized by a prototype feature, and cluster membership is defined by fixed thresholds on geometrical distance (location, log scale) and learned thresholds on appearance descriptor distance [16]. Note that this clustering procedure identifies a large, arbitrary number of clusters, and is capable of robustly grouping features despite imprecision in approximate inter-subject image alignment.

With model features defined, factors in Equation (4) are estimated from feature samples in clusters f_i. Age A is defined as a discrete random variable over K age categories. Gaussian mean and variance parameters of $p(I_j|f_i, A)$ are estimated via maximum likelihood for each age category, probability mass parameters of $p(f_i|A)$ are determined via maximum a-posteriori (MAP) estimation from co-occurrence counts of f_i, A and a Laplace prior. While f_i generally represents a distinctive anatomical structure, a special model feature f_0 is reserved for spurious features arising from background noise. As such features can vary arbitrarily in geometry and appearance, $p(I_j|f_0, A)p(f_0|A)$ is taken to be uniform and constant.

Model fitting is used to interpret image data associated with a new subject in terms of the entire code book F, and to estimate the age or developmental stage. Fitting aims to identify the MAP age estimate A_{MAP} maximizing the posterior probability:

$$A_{MAP} = \underset{A}{\mathrm{argmax}}\{p(A|\bar{I})\}. \tag{5}$$

As in learning, the image is first spatially normalized via subject-atlas alignment. Image features are then extracted and matched to model features, where a match occurs between image and model feature pair (I_j, f_i) if the Euclidean distance between their normalized geometry and appearance descriptors falls within geometry and appearance thresholds associated with f_i. Equation (5) is then evaluated via Equation (4) at each discrete age value to identify A_{MAP}.

3   Experiments

Experiments validate our model on infant brain development in T1-weighted MRI data from the data set of [19], consisting of 92 healthy subjects imaged from one to seven times over an age range of 8-590 days, for a total of 230 images. All subject images are robustly aligned to a single subject arbitrarily selected as an atlas via a similarity transform (rigid + isotropic scaling). Alignment correctness is validated by inspection of resampled images and alignment parameters. Scale-invariant features are extracted in individual aligned images,

after which clustering is used to generate a code book F of model features for analysis. Finally, age-conditional factors $p(I_j|f_i, A)$ and $p(f_i|A)$ are trained based on F and subject age labels A, where A is discretized into 10 bins with equal image counts and approximately equal age ranges. The prior $p(A)$ can be set according to expectations, here it is taken to be uniform.

3.1   Age-Related Structure

Distributions $p(f_i|A)$ quantify the probabilistic relationship between learned anatomical features and age. A variety of criteria could be used to quantify the informativeness of this relationship; we found that information-theoric measures such as entropy tend to overemphasize the importance of spurious, low entropy distributions generated from small numbers of data samples. Here, we consider a Fisher's exact test computed from a 2x2 contingency table of the presence/absence of feature f_i within/without a 2-category (120-day) age window. The null hypothesis is that feature/age are statistically independent, in which case their co-occurrence follows a hypergeometric distribution, which effectively accounts for small sample sizes.

From a total of 6.5K learned model features, 36 are identified with a false discovery rate [20] of 0.05. Figure 1 illustrates the 20 features bearing the most significant relationship with age. Note that significant features show a high degree of symmetry, out of 20 features, eight represent homologous structure identified independently in opposite hemispheres with similar age distributions, another three represent midbrain structure.

3.2   Age Prediction

Model fitting allows age prediction from individual subject images, which may be potentially useful assessing developmental stage. A five-fold cross validation paradigm is adopted, where subjects are randomly divided into 5 mutually exclusive subsets of approximately the same number of images. For each subset, age is predicted for all images based on a model trained on the remaining 4 subsets. Note that different images of the same subject are never present in both testing and training subsets.

The infant brain increases in size, particularly during early weeks, and thus an initial baseline for predicting age from structural MR would be based on volume or size measurements [13]. For comparison, a linear predictor of age based on isotropic subject scale relative to the atlas results in a mean error of 102 days, random guessing results in a mean error of 200 days. Figure 2 illustrates the result of MAP age prediction, which results in a mean error of 72 days. Although this is lower than baseline methods, it is non-negligible, and we hypothesize that error may be related to differences in developmental rates between different subjects. Curves in Figure 2 plot predicted age trajectories for 5 subjects, while a high degree of inter-subject variability is observed, 90% of sequential age predictions follow a monotonically increasing trend. Thus, it may be more accurate to interpret the predicted age as the stage of development
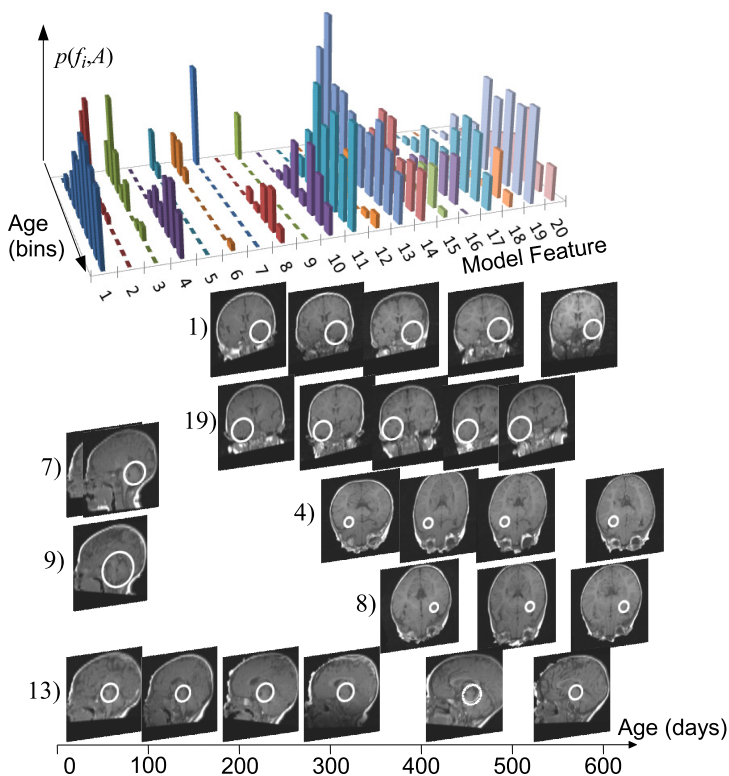
Fig. 1. Top: distribution $p(f_i, A)$ for the 20 most significant age-related features over 10 age categories. Below: visual examples of features (white circles) in subject image slices over age. Pairs (4, 8) and (1,19) represent symmetric white matter patterns appearing at slightly different onsets. (7) and (9) represent distinct modes cerebellar anatomy linked with vermian development and occurring exclusively in early life. (13) occurs in the brain stem across the age range, more frequently in early life. Note the lack of visible white matter under 100 days, e.g. corpus callosum.

with respect to the population, which may generally either lag or lead the actual chronological age; we intend to investigate this possibility in future work.

4   Discussion

This paper presents a novel model describing anatomical development as a collection of features localized over space and time. Validation on infant structural MRI data demonstrates statistically significant age-related features can be automatically discovered across the infant age range, and the first results of infant age prediction are presented. Numerous avenues for future investigation exist, including modeling longitudinal feature-time trajectories or dependencies between features, linking developmental changes other factors such as gender or disease,
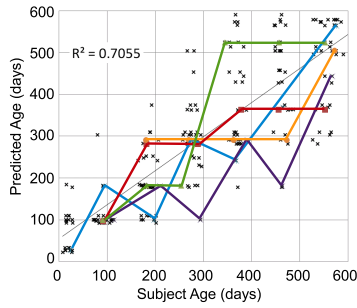
Fig. 2. MAP predicted age vs. subject age across subjects. Thick colored lines illustrate age trajectories for five subjects with scans at 6 or more time points.

incorporating alternative image features such as affine invariant features, investigating alternative means of querying age-related model features. Modeling based on a fixed reference coordinate system here is sufficient for infant brains, an evolving coordinate system [21] will allow application of the modeling to prenatal, fetal and embryonic stages of development.

References

1. Oishi, K., Mori, S., Donohue, P., Ernst, T., Anderson, L., Buchthal, S., Faria, A., Jiang, H., Li, K., Miller, M., van Zijl, P., Chang, L.: Multi-contrast human neonatal brain atlas: Application to normal neonate development analysis. NeuroImage 56, 8–20 (2011)
2. Huppi, P., Warfield, S., Kikinis, R., Barnes, P., Zientara, G.P., Jolesz, F., Tsuji, M., Volpe, J.: Quantitative magnetic resonance imaging of brain development in premature and mature newborns. Ann. Neurol. 43, 224–235 (1998)
3. Woodward, L., Anderson, P., Austin, N., Howard, K., Inder, T.: Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. N. Engl. J. Med. 355, 685–694 (2006)
4. Good, C.D., Johnsrude, I., Ashburner, J., Henson, R., Friston, K.J., Frackowiak, R.S.J.: A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14, 21–36 (2001)

5. Thompson, P.M., Giedd, J.N., Woods, R.P., MacDonald, D., Evans, A.C., Toga, A.W.: Growth patterns in the developing human brain detected using continuum-mechanical tensor mapping. Nature 404(6774), 190–193 (2000)

6. Lao, Z., Shen, D., Xue, Z., Karacali, B., Resnick, S., Davatzikos, C.: Morphological classification of brains via high-dimensional shape transformations and machine learning methods. NeuroImage 21, 46–57 (2004)

7. Franke, K., Ziegler, G., Klöppel, S., Gaser, C.: ADNI: Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: exploring the influence of various parameters. NeuroImage 50(3), 883–892 (2010)

8. Sabuncu, M.R., Van Leemput, K.: The Relevance Voxel Machine (RVoxM): A Bayesian Method for Image-Based Prediction. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 99–106. Springer, Heidelberg (2011)

9. Shi, F., Yap, P., Fan, Y., Gilmore, J., Lin, W., Shen, D.: Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. NeuroImage 51, 684–693 (2010)

10. Paus, T., Collins, D.L., Evans, A.C., Leonard, G., Pike, B., Zijdenbos, A.: Maturation of white matter in the human brain: A review of magnetic resonance studies. Brain Research Bulletin 54(3), 255–266 (2001)

11. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Doria, V., Serag, A., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: A dynamic 4d probabilistic atlas of the developing brain. NeuroImage 54(4), 2750–2763 (2011)

12. Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. NeuroImage 59(3), 2255–2265 (2012)

13. Knickmeyer, R., Gouttard, S., Kang, C., Evans, D., Wilber, D., Smith, J., Hamer, R.M., Lin, W., Gerig, G., Gilmore, J.H.: A Structural MRI Study of Human Brain Development from Birth to 2 Years. Journal of Neuroscience 28(47), 12176–12182 (2008)

14. Lindeberg, T.: Feature detection with automatic scale selection. IJCV 30(2), 79–116 (1998)

15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV 60(2), 91–110 (2004)

16. Toews, M., Wells III, W.M., Collins, D.L., Arbel, T.: Feature-based morphometry: Discovering group-related anatomical patterns. NeuroImage 49(3), 2318–2327 (2010)

17. Mikolajczk, K., Schmid, C.: A performance evaluation of local descriptors. In: CVPR, vol. 2, pp. 257–263 (2003)

18. Comaniciu, D., Meer, P.: Mean Shift: A robust approach toward feature space analysis. IEEE TPAMI 24(5), 603–619 (2002)

19. Almli, C., Rivkin, M., McKinstry, R.: Brain Development Cooperative Group: The NIH MRI study of normal brain development (Objective-2): Newborns, infants, toddlers, and preschoolers. NeuroImage 35(1), 308–325 (2007)

20. Benjamini, Y., Hochberg, Y.: The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188 (2001)

21. Grenander, U., Srivastava, A., Saini, S.: A pattern-theoretic characterization of biological growth. IEEE TMI 26(2), 648–659 (2007)

Constrained Sparse Functional Connectivity Networks for MCI Classification

Chong-Yaw Wee[1], Pew-Thian Yap[1], Daoqiang Zhang[1],
Lihong Wang[2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC,
University of North Carolina at Chapel Hill, NC, USA
[2] Brain Imaging and Analysis Center,
Duke University Medical Center, Durham, NC, USA

Abstract. Mild cognitive impairment (MCI) is difficult to diagnose due to its subtlety. Recent emergence of advanced network analysis techniques utilizing resting-state functional Magnetic Resonance Imaging (rs-fMRI) has made the understanding of neurological disorders more comprehensively at a whole-brain connectivity level. However, inferring effective brain connectivity from fMRI data is a challenging task, particularly when the ultimate goal is to obtain good control-patient classification performance. Incorporating sparsity into connectivity modeling can potentially produce results that are biologically more meaningful since most biologically networks are formed by a relatively few number of connections. However, this constraint, when applied at an individual level, will degrade classification performance due to inter-subject variability. To address this problem, we consider a constrained sparse linear regression model associated with the least absolute shrinkage and selection operator (LASSO). Specifically, we introduced sparsity into brain connectivity via l_1-norm penalization, and ensured consistent non-zero connections across subjects via l_2-norm penalization. Our results demonstrate that the constrained sparse network gives better classification performance than the conventional correlation-based network, indicating its greater sensitivity to early stage brain pathologies.

1 Introduction

Mild cognitive impairment (MCI) is an intermediate stage of brain cognitive decline between normal aging and dementia. MCI is associated with increased risk of developing Alzheimer's disease (AD), especially when memory loss is the predominant symptom. Some individuals with MCI remain stable or return to normal over time, but more than half progress to dementia within 5 years [6]. According to a latest, long-term study of nearly 4000 participants, cognitive impairment has a significant impact on life expectancy similar to chronic conditions such as diabetes or chronic heart failure [14]. Early detection is important for possible delay of the progression of mild MCI to moderate and severe stages. However, diagnosis of MCI is difficult due to its mild symptoms of cognitive impairment, causing most computer-aided diagnosis to achieve lower than desired performance.

Constructing functional brain connectivity from neuroimaging data holds great promise for identifying image-based markers that are important for distinguishing between

MCI and normal aging. A large body of work on functional connectivity modeling has been proposed based on correlation analysis [18, 19]. However, correlation only captures pairwise information and is unable to provide an adequate and complete account of the interaction between many brain regions. Inferring effective brain connectivity from fMRI data for biologically more meaningful interpretation and better classification performance is a challenging task. Many spurious connections arise due to the low frequency (< 0.1Hz) spontaneous fluctuation of blood oxygen level dependent (BOLD) signals and physiological noise such as cardiac and respiratory cycles. Recent work [16] has shown that certain sparsity constraints can be imposed to elucidate robust connections from a set of noisy connections. The sparsity constraint correlates with the fact that, neurologically, a brain region predominantly interacts only with a small number of other regions. Sparse brain connectivity for fMRI data can be constructed through penalizing the linear regression model with l_1-norm as in the least absolute shrinkage and selection operator (LASSO) [10].

However, sparse modeling is unable to deal with inter-subject variability since l_1 penalization at an individual level will result in different network structures across subjects [20], i.e., the non-zero connections are different for each subject. This will inevitably make the comparison between subjects difficult and thus degrade generalization performance of trained classifiers. To address this issue, we propose to employ a constrained sparse linear regression model that minimizes the effect of inter-subject variability in network representation. By this approach, the connection topology is kept identical among subjects, while at the same time allowing individual connection parameters to vary between subjects. This will allow better and more direct comparison among subjects for patient identification. To the best of our knowledge, the current study is the first attempt to construct functional brain network using constrained sparse linear regression model for the purpose of MCI classification. We seek to validate whether this new network modeling strategy can be used to improve classification performance. We will also identify brain regions that contribute most to the classification performance. This paper sheds new light on the effectiveness of applying constrained sparse functional network for diagnosis of progressive neurodegenerative disorders.

2   Materials and Methods

Resting-state fMRI (rs-fMRI) scans of 12 MCI patients and 25 healthy controls were acquired using a 3 Tesla (Signa EXCITE, GE) scanner with the following parameters: TR/TE = 2000/32 ms, flip angle = 77°, imaging matrix = 64×64, FOV = 256×256 mm^2, 34 slices, 150 volumes, and voxel thickness = 4 mm. During scanning, all subjects were instructed to keep their eyes open and stare at a fixation cross in the middle of the screen to prevent them from falling into sleep and the saccade-related activation due to eyes-closed. Informed consent was obtained from all subjects, and the experimental protocols were approved by the institutional ethics board. Confirmation of diagnosis for all subjects was made via expert consensus panels. Demographic and clinical information of the participants is provided in Table 1.

Table 1. Demographic and clinical information of the participants

Group	MCI	Control	p-value
No. of subjects (Male/Female)	6/6	9/16	-
Age (mean $\pm$ SD)	75.0 ± 8.0	72.9 ± 7.9	0.3598^a
Years of education (mean $\pm$ SD)	18.0 ± 4.1	15.8 ± 2.4	0.0491^a
MMSE (mean $\pm$ SD)	28.5 ± 1.5^b	29.3 ± 1.1	0.1201^a

^{a} The p value was obtained by two-sample two-tailed t-test.
^{b} One of the patients does not have a MMSE score.

Post-processing of the fMRI images including slice timing correction and head-motion correction were performed using the Statistical Parametric Mapping (SPM8[1]) software package. The images were then masked with their respective gray matter (GM) masks, created by segmenting the GM regions from their T1-weighted images to eliminate the physiological noise caused by cardiac and respiratory cycles in white matter and cerebrospinal fluid [17]. Then, we parcellated the brain space into 116 ROIs by warping the automatic anatomical labeling (AAL) template to the GM-masked fMRI images. The mean time series of each ROI was computed for each subject. Temporal band-pass filtering of frequency interval ($0.025 \leq f \leq 0.100$Hz) was then performed to minimize the effects of low-frequency drift and high-frequency noise. This frequency interval was further decomposed into 5 equal-length spectral, enabling a more frequency specific analysis of the regional mean time series [19].

2.1 Constrained Sparse Functional Brain Connectivity

Suppose we have N training subjects and M ROIs, the mean time series of p-th ROI for n-th subject, $\mathbf{y}_p^n$, can be regarded as a response vector that is modeled by a linear combination of time courses of other ROIs as

$$\mathbf{y}_p^n = \mathbf{A}_p^n \boldsymbol{\alpha}_p^n + e_p^n, \tag{1}$$

where e_p^n is the noise, $\mathbf{y}_p^n = [y_p^n(1); y_p^n(2); \ldots; y_p^n(T)]$ with T being the number of time points in the time series, $\mathbf{A}_p^n = [\mathbf{y}_1^n, \ldots, \mathbf{y}_{p-1}^n, \mathbf{y}_{p+1}^n, \ldots, \mathbf{y}_M^n]$ is data matrix of the p-th ROI, and $\boldsymbol{\alpha}_p^n = [\alpha_1^n; \ldots; \alpha_{p-1}^n; \alpha_{p+1}^n; \ldots; \alpha_M^n]$ is the weight vector. The sparse brain connectivity modeling of the n-th subject and p-th ROI can be considered as a standard l_1-norm regularized optimization problem with the following objective function

$$f(\boldsymbol{\alpha}_p^n) = \frac{1}{2} \left\| \mathbf{y}_p^n - \mathbf{A}_p^n \boldsymbol{\alpha}_p^n \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_p^n \right\|_1, \tag{2}$$

where $\lambda > 0$ is the regularization parameter controlling the "sparsity" of the model, with a higher value corresponding to a sparser model, i.e., more elements in $\boldsymbol{\alpha}^n$ are zero. Note the l_1-norm penalization is imposed individually on different $\boldsymbol{\alpha}_p^n$ vectors.

For multiple subjects, the objective function in Eq. (2) can be modified as

$$f(\boldsymbol{\alpha}_p) = \sum_{n=1}^N \left(\frac{1}{2} \left\| \mathbf{y}_p^n - \mathbf{A}_p^n \boldsymbol{\alpha}_p^n \right\|_2^2 \right) + \lambda \left\| \boldsymbol{\alpha}_p \right\|_{2,1} \tag{3}$$

---

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

where $\|\boldsymbol{\alpha}_p\|_{2,1}$ is the summation of l_2-norms of non-overlapping $\|\boldsymbol{\alpha}_p^n\|_1$. Specifically, the l_2-norm penalization is imposed on the same elements across different matrices $\boldsymbol{\alpha}_p$ which forces the weights corresponding to certain feature across different subjects to be grouped together. This constraint promotes group-based sparsity by keeping the topology to be identical among subjects, while at the same time allowing variation between subjects. This reduces inter-subject variability and allows for better and more consistent inter-subject comparison for patient identification. The nonzero coefficients in $\boldsymbol{\alpha}$ matrix are treated as an indicator of functional brain connectivity. We use the SLEP toolbox [11] to solve Eq. (3).

2.2   Feature Extraction and Feature Selection

Weighted local clustering coefficient, a measure that quantifies the cliquishness of the nodes, is extracted from all connectivity maps as

$$C_p = \frac{2 \times \sum_{q:q \neq p \in \zeta} e_{p,q}}{k_p(k_p - 1)}, \tag{4}$$

where k_p is the number of ROIs that are connected to the p-th ROI, ζ is the subnetwork comprising nodes directly connected to the p-th ROI, and $e_{p,q}$ is the parameter value between the p-th and q-th ROIs. A total of 116 features are obtained from each map, producing a pool of 580 features for each subject.

After feature extraction, we utilized a hybrid method to select the most relevant features for classification. Two filter-based approaches are initially used to reduce the number of features, followed by a wrapper-based approach to further select a subset of features that is favorable to MCI classification. Specifically, in the first filter-based approach, only those features with p-values smaller than the predefined threshold, measured via between-group t-test, will be retained for subsequent feature selection. Then, the minimum redundancy and maximum relevance (mRMR) algorithm [4] was employed to further exclude redundant features. Finally, the support vector machine (SVM) recursive feature elimination [9], a wrapper-based method, was used to select a subset of most discriminative features for MCI classification.

2.3   Classification

SVM with linear kernel was employed to evaluate the discriminative power of the selected features derived from constrained sparse networks. The optimal SVM models as well as an unbiased estimation of the generalization performance of the complete framework were obtained via a nested cross-validation scheme. For N total number of subjects involved in the study, one was left out for testing, and the remaining $N - 1$ were used for training. From these $N - 1$ samples, $N - 1$ different training subsets were formed by each time leaving one more sample out, giving us $N - 2$ subjects in each training subset. For each training subset, feature extraction and feature selection were performed. The performance of each combination of SVM parameters along with the selected features was evaluated using the second left out subject. The combination that gives the best performance was used to construct the optimal SVM model for future

classification. This procedure was repeated $N - 1$ times, once for each training subset. When the completely unseen (totally left out during the entire training and parameter optimization process) test sample was to be classified, all $N - 1$ classifiers were used, and their outcomes were combined using an averaging operator to provide the final classification decision. This process was repeated N times, each time leaving out a different subject, finally leading to an overall cross-validation classification accuracy. In this study, the optimization of λ parameter in Eq. (3) was performed via grid search.

3   Experimental Results

Constrained sparse connectivity maps of one healthy control and one MCI patient are shown in Figure 1. Spatial connection topology of our sparse networks and the fully-connected correlation-based networks are shown in Figure 2. It can be clearly observed that the generated connectivity networks are significantly sparser than the fully connected correlation-based networks. There are a number of findings that are biologically interesting. First, the bilateral temporal lobes show a relatively smaller amount of intra-lobe connections than other lobes, which has been extensively reported in the literature [16, 18]. Second, there are significantly more inter-lobe connections between parietal and occipital lobes than any other lobe pairs [16], possibly due to compensatory effect. These selected non-zero elements in the constrained sparse matrices reflect the connections that are crucial for discriminating MCI patients from healthy controls.

MCI classification performance of the proposed constrained sparse network was compared with correlation-based connectivity using single and multi-spectral characterization. In the single spectrum case, feature extraction was directly performed on the band-pass filtered BOLD signal without further frequency sub-band decomposition. Sub-band decomposition in the multi-spectral case enables more detailed characterization of subtle changes in BOLD signal, and hence better discriminative power [19]. During evaluation, SVM classifier with the same linear kernel but different hyperparameters was used in a leave-one-out fashion due to the limited number of available samples. MCI classification performance for constrained sparse and correlation-based
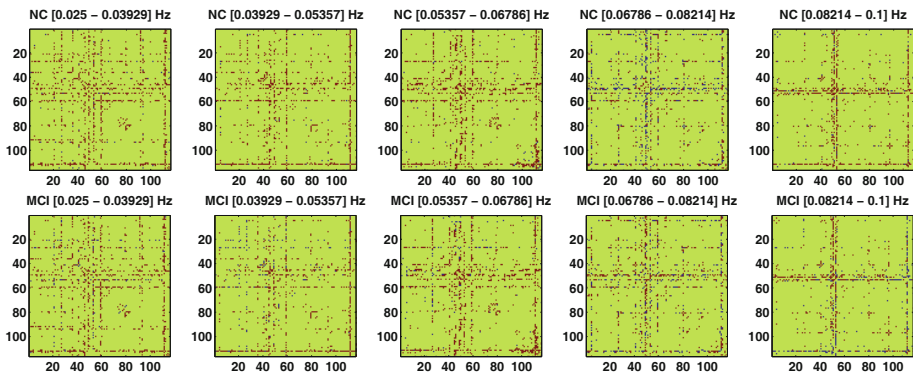

Fig. 1. Constrained sparse connectivity maps with $\lambda = 0.15$. (Red = positive connection, blue = negative connection, green = no connection)
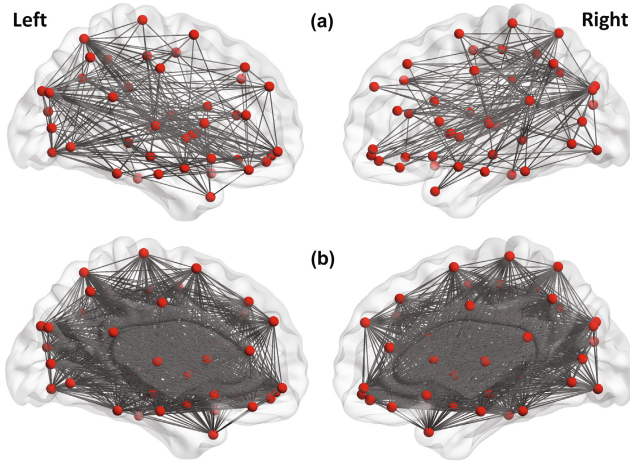
Fig. 2. Constrained sparse networks with $\lambda = 0.15$ (a) and the fully-connected correlation-based networks (b), after excluding the cerebellum

networks are summarized in Table 2. The proposed constrained sparse network with multi-spectral characterization yields the best classification performance with an accuracy of 86.49%, which is an increment of at least 10% from that of the correlation-based approach. A cross-validation estimation of the generalization performance shows an area of 0.8333 under the receiver operating characteristic curve (AUC), indicating good diagnostic power. Note that $\lambda = 0.15$ was found to give the optimal performance.

The most discriminant regions that were selected in the classification process include regions located in frontal lobes (e.g. orbitofrontal cortex [8], frontal gyri [1] and rectus gyrus [5]), temporal lobes (e.g. temporal gyri [2,5,15] and temporal pole [13]), and other regions such as cingulate gyri [7], amygdala [3], angular gyrus [15], and occipital gyri [12], which is in line with the findings that AD, strongly related to episodic memory impairment, causes atrophies in temporal and frontal lobes at the beginning stages of the disease. The selected regions are shown graphically in Figure 3.

Table 2. Classification performance for constrained-sparse and correlation-based networks using single and multi-spectral characterization. (ACC = Accuracy; SEN = SENsitivity; SPE = SPEcificity)

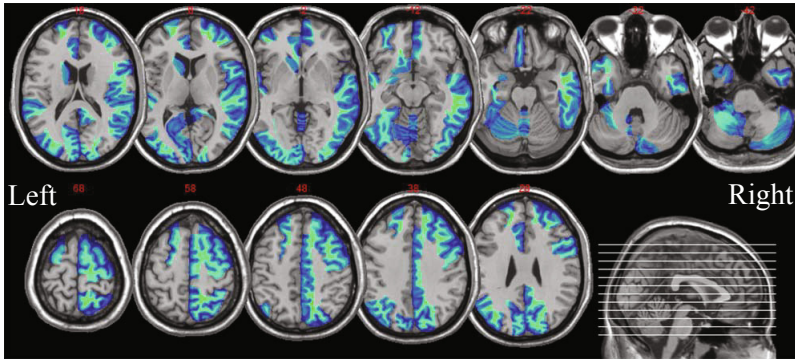| Approach	ACC (%)	AUC	SEN	SPE
Correlation + Single Spectrum	67.57	0.6633	0.0833	0.9600
Sparse + Single Spectrum	72.97	0.6200	0.2500	0.9600
Correlation + Multi-Spectral	75.68	0.7070	0.4167	0.9200
Sparse + Multi-Spectral	86.49	0.8333	0.6667	0.9600

Fig. 3. Most discriminant regions that were selected during MCI classification

4   Discussions and Conclusion

We proposed a novel method to infer functional connectivity networks from rs-fMRI data for the purpose of classification. By imposing group-based sparsity, we minimize spurious connections and inter-subject variability. This is accomplished by considering a constrained sparse linear regression model. Specifically, we incorporate sparsity into brain connectivity estimation via l_1-norm penalization, and ensure inter-subject stability of network structure via l_2-norm penalization. This constrained sparse representation generates topologically consistent functional connectivity networks that allow for better comparison between subjects for classification. The experiment results validate that the proposed method yields markedly improved classification performance compared with the correlation-based network.

References

1. Bell-McGinty, S., Lopez, O.L., Meltzer, C.C., Scanlon, J.M., Whyte, E.M., Dekosky, S.T., Becker, J.T.: Differential cortical atrophy in subgroups of mild cognitive impairment. Arch. Neurol. 62(9), 1393–1397 (2005)
2. Convit, A., de Asis, J., de Leon, M.J., Tarshish, C.Y., De Santi, S., Rusinek, H.: Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer's disease. Neurobiol. Aging 21(1), 19–26 (2000)
3. Dai, W., Lopez, O.L., Carmichael, O.T., Becker, J.T., Kuller, L.H., Gach, H.M.: Mild cognitive impairment and Alzheimer disease: Patterns of altered cerebral blood flow at MR imaging. Radiology 250, 856–866 (2009)
4. Ding, C., Peng, H.: Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 3(2), 185–205 (2005)
5. Fleisher, A.S., Sherzai, A., Taylor, C., Langbaum, J.B., Chen, K., Buxton, R.B.: Resting-state BOLD networks versus task-associated functional mri for distinguishing Alzheimer's disease risk groups. Neuroimage 47(4), 1678–1690 (2009)

6. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J.L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M.C., Whitehouse, P., Winblad, B.: On behalf of the participants of the International Psychogeriatric Association Expert Conference on mild cognitive impairment. Lancet 367, 1262–1270 (2006)

7. Gold, B.T., Jiang, Y., Jicha, G.A., Smith, C.D.: Functional response in ventral temporal cortex differentiates mild cognitive impairment from normal aging. Hum. Brain. Mapp. 31(8), 1249–1259 (2010)

8. Grady, C.L., McIntosh, A.R., Beig, S., Keightley, M.L., Burian, H., Black, S.E.: Evidence from functional neuroimaging of a compensatory prefrontal network in Alzheimer's disease. J. Neurosci. 23(3), 986–993 (2003)

9. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.: Gene selection for cancer classification using support vector machines. Machine Learning 46(1-3), 389–422 (2004)

10. Lee, H., Lee, D.S., Kang, H., Kim, B.N., Chung, M.K.: Sparse brain network recovery under compressed sensing. IEEE Trans. Med. Imaging 30(5), 1154–1165 (2011)

11. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections. Arizona State University (2009), http://www.public.asu.edu/~jye02/Software/SLEP

12. Nobili, F., Mazzei, D., Dessi, B., Morbelli, S., Brugnolo, A., Barbieri, P., Girtler, N., Sambuceti, G., Rodriguez, G., Pagani, M.: Unawareness of memory deficit in amnestic mci: FDG-PET findings. J. Alzheimers Dis. 22(3), 993–1003 (2010)

13. Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S.A., Rodriguez, G., Pagani, M.: Principal component analysis of FDG PET in amnestic MCI. Eur. J. Nucl. Med. Mol. Imaging 35(12), 2191–2202 (2008)

14. Sachs, G.A., Carter, R., Holtz, L.R., Smith, F., Stump, T.E., Tu, W., Callahan, C.M.: Cognitive impairment: An independent predictor of excess mortality: A cohort study. Ann. Intern. Med. 155(5), 300–308 (2011)

15. Smith, C.D., Chebrolu, H., Wekstein, D.R., Schmitt, F.A., Jicha, G.A., Cooper, G., Markesbery, W.R.: Brain structural alterations before mild cognitive impairment. Neurology 68(16), 1268–1273 (2007)

16. Supekar, K., Menon, V., Rubin, D., Musen, M., Greicius, M.D.: Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. PLoS Comput Biol 4, e1000100 (2008)

17. Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L.: Intrinsic functional connectivity as a tool for human connectomics: Theory, properties and optimization. J. Neurophysiol. 103, 297–321 (2010)

18. Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., Jiang, T.: Altered functional connectivity in early Alzheimer's disease: A resting-state fMRI study. Hum. Brain Mapp. 28(10), 967–978 (2007)

19. Wee, C.Y., Yap, P.T., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Resting-state multi-spectrum functional connectivity networks for identification of MCI patients. PLoS ONE 7(5), e37828 (2012)

20. Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. J. Roy. Stat. Soc. B 68(1), 49–67 (2006)

MR-Less Surface-Based Amyloid Estimation by Subject-Specific Atlas Selection and Bayesian Fusion

Luping Zhou[1,3], Olivier Salvado[1], Vincent Dore[1], Pierrick Bourgeat[1],
Parnesh Raniga[1], Victor L. Villemagne[2],
Christopher C. Rowe[2], Jurgen Fripp[1], and the AIBL Research Group

[1] Australian e-Health Research Centre, CSIRO, Australia
[2] Department of Nuclear Medicine and Centre for PET, Austin Hospital, Australia
[3] Faculty of Informatics, University of Wollongong, Australia

Abstract. For clinical evaluation, assessing amyloid deposition with PiB-PET is desirable without requiring MR acquisition and associated fusion/segmentation techniques. A useful clinical tool is to estimate PiB-PET against the brain surface, which is however challenging using PET alone because of the lack of structural information. We propose a method to generate such estimate, where multiple atlases are selected and combined with local weights in a Bayesian framework. Qualitative and quantitative comparison with and without MRI are presented. Using PET only, the average error on the brain surface was around 13% compared to MRI-dependant method.

1 Introduction

Beta-amyloid (Aβ) plaques found in gray matter are among the most prevalent pathological characteristics of Alzheimer's disease (AD), and can be imaged in vivo by Positron Emission Tomography (PET) using ^{11}C-PiB marker. Estimating and visualizing cortical gray matter (GM) Aβ deposition is important for the diagnosis of the disease and the monitoring of plaque reducing therapy.

Because Aβ plaques are present mostly in the gray matter, visualisation on the cortical surface would be a very valuable clinical tool. Such tools are widely used for FDG-PET to assess hypo-metabolism, and a good example is Neurostat [1]. These tools benefit from the fact that the uptake of FDG in the white matter (WM) is very low compared to that of the GM, and therefore a straightforward sum or maximum projection in the GM on the cortical surface is possible. This assumption is not valid for amyloid imaging as most markers, including PiB, have significant retention in the WM, and therefore the 3D surface separating WM and GM is not well-defined and has to be estimated. This is very challenging as PET imaging lacks the anatomical details and the resolution to resolve this surface. As a result, Magnetic Resonance Imaging (MRI) is required to segment the brain into GM and WM, and extract the separating surface. This surface is then registered to the PET scan where GM PiB retention can be measured and displayed.

In standard clinical setup, when PET amyloid imaging is visually assessed, MRI scans are not always available (e.g. different system, posterior scanning time) and in a substantial number of cases not possible (e.g. presence of metallic implant). This paper proposes a method to estimate PiB-PET on the cortical surface without the need of an MRI scan. Our proposed "PET-only" approach is compared to the best case scenario when an MRI is available, which we refer to as the MRI-dependent method.

A PET-only method was reported in [2]. It employed a single MRI atlas with segmented tissues and co-registered to a PET image from the same subject. When a new subject PET was registered to the atlas PET, the atlas MRI could be used to estimate the subject GM. The maximal PiB uptake within the subject GM was measured along the normal direction of the brain surface. Although no quantitative precision analysis of this method is reported, our experiments found that its estimation accuracy highly depends on the atlas selection due to i) different anatomy between the single atlas and the subject, and ii) using hard PET tissue partitions simply copied from the single warped MRI atlas.

In this paper we address these problems by introducing a robust PET-only method that significantly improved upon [2]. In particular we achieved this by using 1) multiple possible GM-WM surfaces, 2) multiple likely GM tissue maps, 3) local and adaptive atlas selection customized for each subject and 4) a probabilistic framework to combine the estimates from individual atlases. Our method was evaluated both visually and quantitatively. Our experiments showed that the estimate without MRI is about 13% different from that with MRI at each vertex, and the intra-class correlation between these two methods is reached 0.94.

2   Method

2.1   Overview

Our proposed method directly estimates GM PiB-PET retention without an explicit GM segmentation by using multiple atlases.

Twenty subjects (with both PET and MRI) were used as atlases, whose MRI images were co-registered and segmented into GM, WM and CSF tissue maps. These atlases were randomly selected to match the full range of disease progression and are reasonably representative. The 3D GM-WM surfaces of the atlases were extracted and registered to a canonical space (arbitrary subject) to obtain vertex correspondence (Section 2.2). For each new subject without MRI, the subject PET image was registered with the twenty atlas PET images (Section 2.3), allowing the subject PiB to be estimated within the atlas GM tissue maps independently. Finally, the most similar atlases for each subject were selected by local and adaptive weighting (Section 2.4), and combined using a Bayesian framework (Section 2.5) to generate the subject's unique GM PiB estimate.

2.2   Surface Registration

To attain vertex correspondence, all atlas surfaces were registered using a multi-scale EM-ICP [3]. Briefly, surface curvatures were obtained by geometric surface

simplifications at different scales to reveal local structures. By controlling the degree of surface simplification, only major sulci and gyri common to all the subjects were kept for EM-ICP registration, while the subject-specific structures were ignored. After registration, the atlas surfaces were resampled with the same number of vertices (80,000).

2.3 Affine Registration for PET Images

The subject and atlas PET images were co-registered to transform i) the 20 atlas surfaces and ii) the 20 sets of the tissue probability maps to the subject space. Affine registration based on block matching of feature points in PET images was employed [4] instead of non-rigid registration because of the limited resolution and noise in the PET.

2.4 Local Subject-Specific Atlas Selection

For a given subject, not all 20 atlases were suitable for its GM PiB estimation. Some atlases could be very different due to the variation of individual brain shapes and the $A\beta$ deposition that could be affected by atrophy. Therefore, an optimal subset of atlases was selected for each subject. The selection number is set as 10, which was experimentally found a suitable choice.

Our atlas selection was local and adaptive. "Local" means the selection of atlases was determined by a local metric computed between small PET image blocks. "Adaptive" means the selected subset of atlases was different from vertex to vertex. Our atlas selection process was elaborated as follows. For a given surface vertex, the PET image similarity between the subject and an atlas was assessed in a $30 \times 30 \times 30$ (voxels) neighborhood by normalized mutual information (NMI) [5]. The neighborhood size should not be set too small to avoid overfitting noise due to the low resolution of PET. The ten most similar atlas PET images were selected to generate the final estimation at each vertex. We found that local weighting provided superior results over global weighting. Moreover, keeping only a subset of the most similar atlases can further improve the scenario where all atlases are used [6]. Our local adaptive atlas selection reduces errors due to mismatches in the distribution of the plaques and GM shapes, allowing it to successfully handle more unusual cases.

2.5 Surface-Based Measurement by Atlas Fusion

We propose a Bayesian fusion framework for PiB quantification by multi-atlases.

Given a PET image $\mathbf{I}(x)$, where x denotes an image voxel, our target is to measure the mean PIB uptake in GM along the normal directions of the transformed atlas surface $\mathbf{S}^T$. That equals to estimate the expectation $\underset{x \in \Delta}{\mathbb{E}}[\delta(\mathbf{I}, x, l)]$, where $\delta(\mathbf{I}, x, l)$ is an indicator function:

$$\delta(\mathbf{I}, x, l) = \begin{cases} \mathbf{I}(x), & for \; l = 1 \\ 0, & elsewhere, \end{cases}$$

and Δ denotes the line where the surface normal direction meets the image. The symbol l is the tissue label at the voxel x, representing GM, WM and CSF with the values of 1, 2 and 3, respectively. Taking discrete x, we have

$$\mathbb{E}[\delta(x, l)] = \sum_{x \in \Delta} \delta(\mathbf{I}, x, l) P(\mathbf{I}, x, l) dx$$
$$= \sum_{x \in \Delta} \mathbf{I}(x) P(l = 1|\mathbf{I}, x) P(\mathbf{I}, x) dx. \quad (1)$$

Assuming that x is evenly sampled from Δ, the probability $P(\mathbf{I}, x) = 1/|\Delta|$, where $|\Delta|$ is the length of Δ. The posterior label probability $P(l|\mathbf{I}, x)$ is estimated from the transformed atlases $\mathbf{A}_i^T$ ($i = 1 \cdots n$, n is the number of atlases) by maginalizing the joint probability $P(l, \mathbf{A}_i^T|\mathbf{I}, x)$:

$$P(l|\mathbf{I}, x) = \sum_{i=1}^{n} P(l, \mathbf{A}_i^T|\mathbf{I}, x) = \sum_{i=1}^{n} P(l|\mathbf{A}_i^T, \mathbf{I}, x) P(\mathbf{A}_i^T|\mathbf{I}, x). \quad (2)$$

Here $P(l|\mathbf{A}_i^T, \mathbf{I}, x)$ represents the GM probability at the voxel x in the atlas $\mathbf{A}_i^T$, which was obtained in our case from the transformed atlas tissue maps. The term $P(\mathbf{A}_i^T|\mathbf{I}, x)$ is the probability to be well aligned at the voxel x between the target image $\mathbf{I}$ and the transformed atlas $\mathbf{A}_i^T$. In our approach, $P(\mathbf{A}_i^T|\mathbf{I}, x)$ was set proportional to the reciprocal of the NMI metric estimated locally within the neighborhood $N(x)$ of x, i.e., $P(\mathbf{A}_i^T|\mathbf{I}, x) = P(\mathbf{A}_i^T|\mathbf{I}, N(x))$. As aforementioned, the size of $N(x)$ should be reasonably set to avoid overfitting noise. In our approach, $N(x)$ covers all likely GM voxels along the line Δ. Thus, $P(\mathbf{A}_i^T|\mathbf{I}, N(x))$ is constant to the variable x ($x \in \Delta$). Combining (1) and (2), we have

$$\mathbb{E}[\delta(x, l)] = \frac{1}{|\Delta|} \sum_{x \in \Delta} \mathbf{I}(x) P(l = 1|\mathbf{I}, x)$$
$$= \frac{1}{|\Delta|} \sum_{x \in \Delta} \mathbf{I}(x) \sum_{i=1}^{n} P(l = 1|\mathbf{A}_i^T, \mathbf{I}, x) P(\mathbf{A}_i^T|\mathbf{I}, N(x))$$
$$= \sum_{i=1}^{n} P(\mathbf{A}_i^T|\mathbf{I}, N(x)) \left[\frac{1}{|\Delta|} \sum_{x \in \Delta} \mathbf{I}(x) P(l = 1|\mathbf{A}_i^T, \mathbf{I}, x) \right] \quad (3)$$

Eqn. (3) shows an additive property: the estimation from multiple atlases can be attained from each single atlas independently and then linearly combined by local weights. Such a combination is nonlinear for the whole surface. This additive property is important because it allows to refine the estimate with new atlases without recomputing the whole set, thus being computationally efficient.

When estimating the PiB value $\mathbf{I}(x)$ from a single atlas $P(\mathbf{A}_i^T|\mathbf{I}, N(x))$, the weights were the GM probability at x. This implicitly defined a gray matter region with a soft boundary, which reflected the variation observed in the training population. Unlike [2] where hard GM segmentation in a single atlas was used, we found our approach more robust to registration error between PET and MRI.

2.6   SUVR Normalization

In some applications, to ensure inter-subject comparisons, the PiB-PET uptake values need to be normalized by standard uptake value ratio (SUVR) [7]. SUVR is defined as the value of a region containing specific binding to one without specific binding (such as cerebellar gray matter). In such cases, the intensity values $\mathbf{I}(x)$ in the original PET image are homogeneously scaled by a parameter k. As NMI matches structures instead of intensity values, $P(\mathbf{A}_i^T|\mathbf{I}, N(x))$ is invariant to k, and so is the GM probability $P(l = 1|\mathbf{A}_i^T, \mathbf{I}, x)$. Thus,

$$\mathbb{E}[\delta(x,l)] = k \sum_{i=1}^{n} P(\mathbf{A}_i^T|\mathbf{I}, N(x)) \left[\frac{1}{|\Delta|} \sum_{x \in \Delta} \mathbf{I}(x) P(l = 1|\mathbf{A}_i^T, \mathbf{I}, x) \right], \quad (4)$$

where k is determined by SUVR. Eqn. (4) shows that when SUVR changes, our estimation (3) can be simply scaled by k.

PiB-only SUVR normalization is beyond the scope of this paper and several methods have been proposed [8]. In the following, we set k = 1 to focus on evaluating the pure performance of our method by comparing surface PiB estimation with and without MRI.

2.7   Validation

The proposed method was validated on 143 subjects from the Australia Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL). Please note that, the validation on our method does not require SUVR normalization that can be computed without MRI. We nonetheless preprocessed SUVR using MRI cerebellum just for the purpose of clinically relevant analysis (grouping subjects) and display (Fig. 1). With a cutoff value of 1.5, these subjects were categorized into 73 PiB$^+$ and 70 PiB$^-$ subjects.

To compare our method with the best case of senario, MRI-dependent method was applied as ground truth. The estimation difference VAR (%) between the two methods was measured in ratios and averaged over all subjects. It was computed at both ROI and vertex levels. In addition to VAR, we also computed the Pearson correlation and the intra-class correlation (ICC) between the two methods by correlating the estimations at each vertex.

3   Experimental Results and Discussion

We compared the performance of the proposed method with i) the MRI-dependent method (ground truth) and ii) the single-atlas based PiB-only method.

3.1   Comparison with MRI-dependent Method

In Fig. 1, four typical examples are shown. MRI-dependent (top row) and our PET-only (bottom row) methods present similar patterns and are visually very
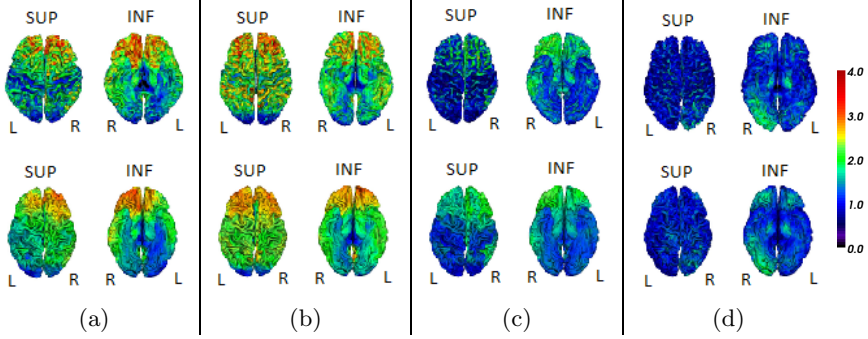
Fig. 1. Surface-based PIB measurements from the MRI-dependent method (the top row) and the proposed method (the bottom row) for four examples: (a) PiB^+ AD, (b) PiB^+ NC, (c) PiB^+ NC, and (d) PiB^- NC.

close, for example, the asymmetric patterns in the left and right hemispheres for subjects (a) and (d).

Table 1 summarizes the quantitative results between the MRI-dependent and the PET-only methods. As shown, the PiB^+ group had lower VAR ($2.2\pm1.5\%$ for ROI, $13.1\pm1.5\%$ for vertex) than the PiB^- group ($3.6\pm2.3\%$ for ROI, $16.4\pm2.1\%$ for vertex), as well as attaining higher Pearson correlation / ICC (0.75 / 0.94) than the PiB- group (0.42 / 0.73). This difference is expected, because the PiB^- group has minimal PiB retention, whose image signal is proportionally overwhelmed by noise than that of the PiB^+ group. The high estimation accuracy of the PiB^+ group in our method is promising as the PiB^+ group is clinically more interested due to its association with possible AD. Meanwhile, the estimate accuracy for the PiB^- group was sufficient for identifying visually similar patterns to the MRI-dependent method (Fig. 1 (d)). We also found that our estimation errors for both groups were close to the reported reproducible errors of PiB quantification using only 30 min imaging [9] (cited in Table 1), which reinforced the precision of our method. Finally, although PET-based SUVR estimation was not included in our method, we also tested the overall performance with the SUVR method in [8]. It showed that the SUVR method would increase the error per vertex by approximately 4% for both groups.

3.2 Comparison with Single Atlas Approach

To demonstrate the advantage of using multiple atlases, the VAR and the correlations with the MRI-dependent method were compared between 10 randomly selected single-atlas and the proposed multiple-atlas approach. The VAR and correlations were compared subject by subject (averaged over all vertices within each subject, Fig. 2 (a) and (b)), and ROI by ROI (averaged over PiB^+ subjects[1], Fig. 2 (c) and (d)). In Fig. 2, the red line was the result from the multiple-atlas approach, and the ten other lines corresponded to the ten single-atlas in comparison. As shown, the proposed approach exhibited significantly lower average VAR

---

[1] Improvement on PiB^- is similar to that of PiB^+.

Table 1. Comparison of estimations from the MRI-dependent and PET-only methods

	PiB$^+$	All	Frontal	Occipital	Temporal	Parietal	Post-cingulate	Putamen
ROI	VAR (%)	2.2 ± 1.5	2.3 ± 1.6	2.1 ± 1.5	2.1 ± 1.4	2.1 ± 1.5	2.4 ± 1.7	2.0 ± 1.4
	cited VAR (%)	-	3.9	3.7	3.3	3.7	4.9	5.1
Ver-tex	VAR (%)	13.1 ± 1.5	13.2 ± 1.6	13.1 ± 1.5	12.9 ± 1.5	13.2 ± 1.5	13.2 ± 1.6	13.3 ± 1.7
	P-Corr (R)	0.75	0.74	0.75	0.76	0.75	0.75	0.75
	ICC	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	cited ICC	-	0.95	0.94	0.97	0.94	0.88	0.89
	PiB$^-$	All	Frontal	Occipital	Temporal	Parietal	Post-cingulate	Putamen
ROI	VAR (%)	3.6 ± 2.3	3.6 ± 2.4	3.4 ± 2.1	3.5 ± 2.2	3.5 ± 2.4	3.2 ± 2.3	3.8 ± 2.4
	cited VAR (%)	-	2.7	3.2	2.5	2.0	0.9	4.1
Ver-tex	VAR (%)	16.4 ± 2.1	16.4 ± 2.1	16.3 ± 2.0	16.2 ± 2.0	16.6 ± 2.2	16.4 ± 2.2	16.6 ± 2.1
	P-Corr (R)	0.42	0.42	0.43	0.42	0.41	0.43	0.41
	ICC	0.72	0.73	0.73	0.72	0.72	0.76	0.68
	cited ICC	-	0.73	0.75	0.69	0.59	0.75	0.66

(a)   (b)

(c)   (d)
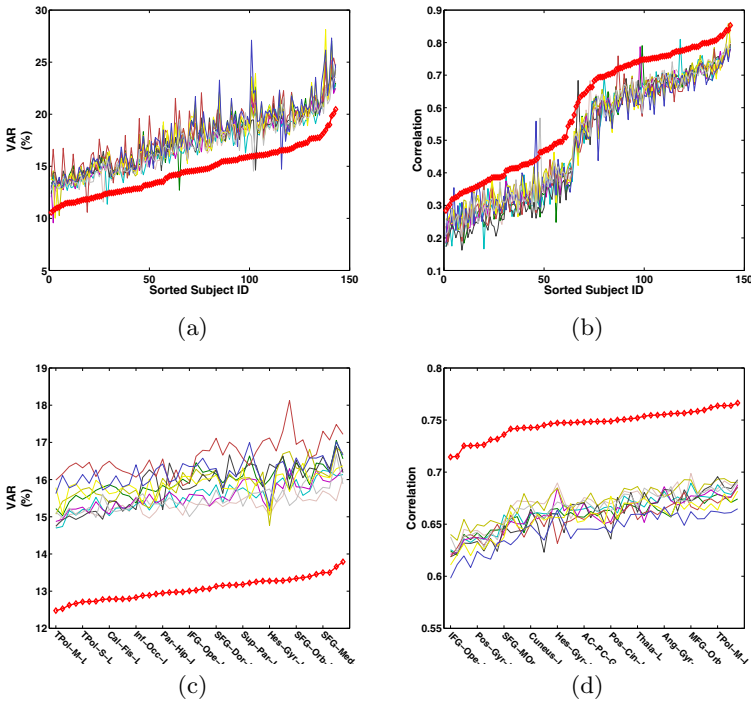
Fig. 2. Comparison of the proposed multi-atlas approach with ten single-atlas approaches by error ratios for (a) each subject and (c) each ROI, and by correlations for (b) each subject and (d) each ROI

and higher correlations over almost all subjects than any single atlas produced. Consistently, when breaking the estimates into surface ROIs and averaged over subjects, the proposed approach revealed an even more salient advantage than any single-atlas in all the ROIs.

4   Conclusion

In this paper, we propose to estimate amyloid deposition against brain surfaces utilizing only PiB-PET images without MRI. The demonstrated accuracy of our method suggests it could provide a CAD tool for evaluation of amyloid imaging. In future, we will extend our approach to include the PiB-PET based SUVR estimation.

References

1. NeuroStat/3D-SSP, http://128.95.65.28/~Download
2. Lilja, A., Thurfjell, L.: Tools for aiding in the diagnosis of neurodegenerative diseases (2010), http://www.faqs.org/patents/app/20100080432
3. Dore, V., Fripp, J., Bourgeat, P., Shen, K., Salvado, O.: Surface-base approach using a multi-scale em-icp registration for statistical population analysis. In: DICTA (2011)
4. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3d structure from serial histological sections. Image Vis. Comput. 19(1), 25–31 (2001)
5. Studholme, C., Hill, D., Hawkes, D.: An overlap invariant entropy measure of 3d medical image alignment. Pattern Recogn. 32(1), 71–86 (1999)
6. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination strategies in multi-atlas image segmentation: application to brain mr data. Trans. Med. Imaging 28(8), 1266–1277 (2009)
7. Lopresti, B., Klunk, W., Mathis, C., Hoge, J., Ziolko, S., Lu, X., Meltzer, C., et al.: Simplified quantification of pittsburgh compound b amyloid imaging pet studies: a comparative analysis. J. Nucl. Med. 46(12), 1959–1972 (2005)
8. Raniga, P., Bourgeat, P., Villemagne, V., O'Keefe, G., Rowe, C., Ourselin, S.: Spline Based Inhomogeneity Correction for 11C-PIB PET Segmentation Using Expectation Maximization. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 228–235. Springer, Heidelberg (2007)
9. Aalto, S., Scheinin, M.N., Kemppainen, M.N., Någren, K., et al.: Reproducibility of automated simplified voxel-based analysis of pet amyloid ligand [11c]pib uptake using 30-min scanning data. Eur. J. Nucl. Med. Mol. Imaging 36, 1651–1660 (2009)

Hierarchical Structural Mapping for Globally Optimized Estimation of Functional Networks

Alex D. Leow[1,2,3,*] Liang Zhan[4,*], Donatello Arienzo[5],
Johnson J. GadElkarim[1,6], Aifeng F. Zhang[1], Olusola Ajilore[1],
Anand Kumar[1], Paul M. Thompson[4], and Jamie D. Feusner[5]

[1] Department of Psychiatry, University of Illinois, Chicago, IL, USA
[2] Department of Bioengineering, University of Illinois, Chicago, IL, USA
[3] Community Psychiatry Associates, Sacramento, CA, USA
[4] Laboratory of Neuro Imaging, Department of Neurology, UCLA, CA, USA
[5] UCLA Semel Institute for Neuroscience and Human Behavior, UCLA, CA, USA
[6] Department of Electrical and Computer Engineering, UIC, IL, USA
alexfeuillet@gmail.com, lzhan@loni.ucla.edu

Abstract. In this study, we propose a framework to map functional MRI (fMRI) activation signals using DTI-tractography. This framework, which we term functional by structural hierarchical (FSH) mapping, models the regional origin of fMRI brain activation to construct "N-step reachable structural maps". Linear combinations of these N-step reachable maps are then used to predict the observed fMRI signals. Additionally, we constructed a utilization matrix, which numerically estimates whether the inclusion of a specific structural connection better predicts fMRI, using simulated annealing. We applied this framework to a visual fMRI task in a sample of body dysmorphic disorder (BDD) subjects and comparable healthy controls. Group differences were inferred by comparing the observed utilization differences against 10,000 permutations under the null hypothesis. Results revealed that BDD subjects under-utilized several key local connections in the visual system, which may help explain previously reported fMRI findings and further elucidate the underlying pathophysiology of BDD.

Keywords: DTI, HARDI, fMRI, network, Simulated Annealing.

1 Introduction

How the brain is organized into functional networks remains an elusive question. Despite widespread use of functional neuroimaging, inferences on network properties have largely been based on associative activation patterns; its limited temporal resolution hinders an understanding of dynamic interactions between spatially defined nodes. Advanced modeling techniques such as Dynamic Causal Modeling [1] provide estimates of complex neuronal dynamic interactions estimated from BOLD signal patterns, but may be practical for understanding only a small number of nodes.

Inferences about network dynamical interactions on a larger scale could potentially be derived using strategies involving structural to functional mapping. This is based on

---

the assumption that a system's structural connectivity pattern provides a framework for determining information available as inputs from certain regions and its outputs and potential influence on other regions [2]. Such an approach could utilize a combination of DTI-tractography to estimate brain white matter connectivity and functional magnetic resonance imaging (fMRI) to estimate the neuronal activity coupled to blood flow changes in anatomical regions that comprise nodes of the network.

Several structural to functional mapping approaches have been proposed [2-9]. Some have focused on specific, limited regional activation patterns [2, 3] while others have developed models for understanding functional connections in larger sets of regions comprising networks or systems [4-11]. One study simultaneously modeled DTI-tractography and fMRI in individuals in the resting state [12]; however, to our knowledge no study has used structural to functional mapping to estimate information transfer on a system-wide level using task-based fMRI data. We therefore developed a model to parsimoniously explain overall observed activation patterns with the assumption that the system is comprised of a set of interacting nodes.

To test this approach, we applied it to understanding activation patterns within the visual system. The visual system has been well characterized in monkeys and humans and can be reliably activated with functional neuroimaging paradigms. In addition, previous studies have provided models for a temporal sequence of activation, particularly for the "first wave" of information transfer (see, for example [13, 14]), from V1 to V2 and other visual cortical areas, followed by temporal and parietal regions. This allows us to make several simplifying assumptions about direction of information transfer. We tested our model in a dataset consisting of healthy human controls and individuals with body dysmorphic disorder (BDD). Individuals with BDD have perceptual distortions in which they misperceived defects in their appearances [15]. Previous fMRI studies in BDD found abnormal visual processing in primary and secondary visual processing systems for face and inanimate object (house) stimuli [16-18]. To better characterize these findings, we thus developed a strategy for functional by structural hierarchical mapping to estimate step-wise activation patterns in the visual system. Here, the word "hierarchical" refers to the n-step structural maps used for model fitting. The fMRI portion of this dataset, which utilized well-characterized house visual stimuli, was previously analyzed for group differences using the general linear model [16]. Here we demonstrate a model of hierarchical structural to functional mapping that, under several simplifying assumptions, elaborates on previous results by providing estimates of structural connection utilization within the visual system.

2   Methods

2.1   Data Collection

Eleven unmedicated participants with BDD and 13 healthy controls, aged 20 to 48 years, were scanned in a 3-T Siemens Allegra MRI scanner. Diffusion-weighted MR imaging (DWI) data were acquired using single-shot spin-echo echo-planar imaging (EPI) (field of view=240mm; voxel size=2.5x2.5x3.0mm, with 0.75 mm gap; TR/TE=7400/96ms; flip angle 9°). We collected 44 contiguous axial slices aligned to

the AC-PC line along 34 gradient-sensitizing directions with b=1000s/mm^2 and one minimally diffusion-weighted scan. DWI data were corrected for eddy current artifacts using FSL toolbox (http://www.fmrib.ox.ac.uk/fsl/fdt/fdt_eddy.html). fMRI data came from a previous object-processing fMRI study; subject selection, acquisition and preprocessing are as previously described [16]. fMRI stimuli consisted of photographs of houses that contained only low spatial frequency (LSF) information [19], in order to more sensitively probe configural and holistic visual processing.

2.2   Brain Network Computation

We computed whole-brain deterministic DTI tractography using Diffusion Toolkit with the Fiber Assignment by Continuous Tracking (FACT) algorithm [17] and a maximum turn angle of 35 degrees. Cortical regions of interest (ROIs) were defined using the Harvard Oxford Cortical probabilistic atlas. The masks were set to a liberal threshold of 10% to allow for the inclusion of tissue along the gray-white matter interface, where DTI tractography estimates are most reliable [21]. To register these ROIs to each subject's DTI space, we used FSL's FLIRT program [22] to determine the optimal affine transformation between the MNI152 T1 average brain (in which the Harvard Oxford probabilistic atlases are based) and each subject's unique FA image, using a mutual information-based cost function. We applied the resulting transformation to register the 26 visual pathway related ROIs to each subject's DTI space. For each pair of ROIs, the number of fibers connecting them was counted to construct a structural connectivity matrix (a fiber was considered to connect two ROIs if it originated in the first ROI and terminated in the second, or vice versa). Thus, this matrix was symmetric, with the diagonal entries set to zero (no self connections).

To generate functional MRI signals for each region of interest, we computed the percentage signal change using the Featquery tool in FSL to estimate BOLD signal differences between the house stimuli and baseline.

2.3   Functional by Structural Hierarchical (FSH) Mapping

Several assumptions and simplifications are needed in order to perform FSH mapping, outlined step-by-step as follows:

1. A collection of cortical anatomical regions or "nodes" is determined to be activated during the fMRI task, and cortical fMRI activation is assumed to originate in a specific region(s) (the "origin"), in this case the left and right intracalcarine cortices.
2. N-step reachable structural maps, defined as regions reachable from the origin at the n-th step of structural connections, are computed by tracing the anatomical connectivity (according to the structural connectivity matrix C) from the origin.
3. Nodes that are reached in the first n steps will be excluded from later steps. All connections between nodes are hypothesized to be excitatory. Moreover, feedback or interactions between nodes that have been reached is assumed to be numerically negligible.

4. The presence of an edge connecting any node pair in the structural connectivity matrix predicts the existence of neuroanatomical connections between regions, which may are may not be "utilized" in the particular fMRI task of interest. A connection between node i and j is considered "utilized" if including the anatomical connection between node i and node j better predicts the overall fMRI activation. This is modeled by a binary utilization matrix U (i.e., if U(i,j)=1, then the structural connection between nodes i and j are utilized in the fMRI task; zero otherwise)

5. The observed fMRI activation pattern is assumed to be explained by linear combinations of n-step reachable structural maps.

Now, we illustrate how to set up the FSH mapping using the BDD house processing example. In step 1, a total of 26 visual system ROIs were carefully selected from the group mean fMRI activation maps overlaid on the Harvard-Oxford cortical probabilistic atlas for house stimuli viewing vs. baseline, for combined BDD and healthy control groups. These 26 regions are the (bilateral) lateral occipital cortex superior division (nodes 1,2), lateral occipital cortex inferior division (3,4), intracalcarine cortex (5,6), precuneus cortex (7,8), cuneal cortex (9,10), parahippocampal gyrus posterior division (11,12), lingual gyrus (13,14), temporal fusiform cortex anterior division (15,16), temporal fusiform cortex posterior division (17,18), temporal occipital fusiform cortex (19,20), occipital fusiform cortex (21,22), supracalcarine cortex (23,24), and occipital pole (25,26).

As the fMRI task was a visual task, we thus assumed that cortical fMRI signals originated in the bilateral primary visual cortex (V1 or intracalcarine cortex; nodes 5 and 6). Mathematically, this is represented by a 26-element "activation" column vector A, whose entries are all zeros except for bilateral V1 (where A takes values of 1).

In steps 2-4, we first observe that 0-step reachable map is A itself, and define the 1-step reachable map as $(U \circ C) \cdot A$ (here the circle operator denotes the entry-wise or Hadamard product of two matrices of the same dimension). The n-step reachable structural maps (A_n) were then determined by the iterative application of the following procedure:

$A_0 = A,\ C_o = C;\ for\ n = 0,\ 1,\ ...,$

1. *initialize* $C_{n+1} = C_n$;

2. *compute the column vector* $A_{n+1} = (U \circ C_n) \cdot A_n$; *for all i such that* $A_{n+1}(i) > 0$, *set all entries in the* $i-th$ *column and* $i-th$ *row of* C_{n+1} *to zero.*

Lastly, we fit the following linear equation so that the entire set of observed fMRI activations in the visual system, S, can be explained by linear combinations of A_n

$$S = w_0 A_0 + w_1 A_1 + w_2 A_2 + ... w_n A_n + e$$

Here, S is the fMRI signal column vector (26x1) and e is the error term. The optimization of the FSH mapping is achieved in two parts. First, U was initialized as a 26x26 matrix whose every entry took a value of 1 (i.e., all structural connections were utilized). We then fixed U and the linear equation above was fitted to determine the weights (w_0, w_1, w_2, etc in the above equation) and to calculate the fitting errors. For the second part of the optimization, the weights were fixed, and data from all subjects in each group were pooled together to estimate a group-level optimal U using

the simulated annealing algorithm (thus yielding two U matrices, one for the BDD and one for the control group). To this end, we perturbed U (from state i to state j) by randomly picking one element in U and changing its value (between 0 and 1). The acceptance criterion determined whether the new state j was accepted from the current state i by applying the following decision rule with respect to an artificial cooling temperature (c).

$$P_c\{accept\ j\} = \begin{cases} 1 & if\ e(j) \le e(i) \\ \exp\left(\dfrac{e(i) - e(j)}{c}\right) & if\ e(j) > e(i) \end{cases}$$

Where e(i) and e(j) are the linear fitting residuals. This perturbation was repeated numerous times until the solution space was adequately sampled and the global minimum reached.

2.4 Determining Statistical Significance for Observed Utilization Group Differences

As structural connection strengths vary across edges, utilization differences for connections that have on average higher fiber counts should be considered potentially more significant. As a result, we devised a statistic dU, by summing up all rows and columns of the difference matrix (subtracting the U of the control group from that of the BDD) weighted by the mean connectivity matrix $\overline{C}$ (structural connectivity matrices averaged across all study participants).

$$dU = \sum_{i,\,j=1}^{26}\left[\left(U^{BDD} - U^{control}\right)\mathbf{o}\overline{C}\right]$$

Here, dU measures the overall group utilization differences, and permutation testing can be conducted to determine whether the observed dU reaches statistical significance. To this end, we performed 10,000 permutations by randomly re-assigning each subject's diagnosis (into two groups of size 11 and 13). At each permutation, the test statistic dU was re-computed and recorded, and the observed dU was then ranked against the re-sampled dUs. If the observed dU was relatively extreme and ranked among the top 5%, then we determined that the observed utilization group difference has reached statistical significance at a p value of 0.05.

3   Results

Linear fitting errors vs. the level of n-step structural maps used for model fitting for one subject is shown in Figure 1. Figure 2 summarizes fitting results for all subjects in both groups. Linear combinations of 1-, 2-, and 3-step structural maps were tested. Three-level structural maps provided sufficient linear fitting, which was further improved with simulated annealing.

Group differences in the binary utilization matrices (BDD-control) weighted by the mean connectivity strength (note that this matrix is symmetrical, and thus its (i, j) and

(j,i) elements indicate the utilization difference for the same structural connection) are shown in Figure 3. According to this model, the control group utilized several more structural connections than the BDD group (the most significant differences, marked A, B, and C in Figure 3, are structural connections between the left intracalcarine cortex and left lingual gyrus, the right intracalcarine cortex and right lingual gyrus, and the right temporal occipital fusiform cortex and right temporal fusiform cortex, posterior division. These connections are visualized in Figure 4. Permutation tests confirmed that the observed utilization group difference reached statistical significance (p = 0.0329).
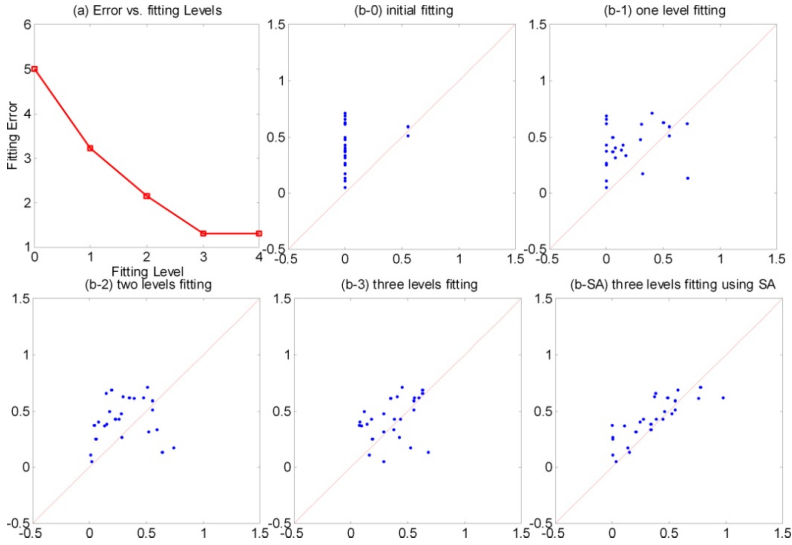

Fig. 1. (a) Fitting errors vs. fitting levels in one study participant. (b-0 through b-3): functional by structural hierarchical (FSH) mapping results in the same subject using 0, 1, 2, and 3-level structural maps. The y axis indicates observed fMRI % signal change and the x axis the predicted fMRI % signal change (data points that are perfectly predicted by the proposed method would lie along the x=y line). (b-SA): the 3-level linear fitting in addition to the simulated annealing (SA) step. Visually, the SA step further improves data fitting.
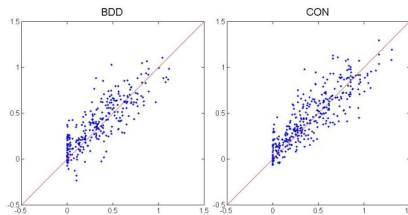

Fig. 2. Group fitting. This shows the fitting of data points, pooled from all subjects in each group, using the proposed 3-level fitting with simulated annealing. The y axis indicates observed fMRI % signal change and the x axis the predicted fMRI % signal change. (Note that predicted values were set to be greater than zero as one of the simplifying assumptions of the model).

4   Discussions and Conclusions

These results are consistent with previous fMRI results for house processing, which found relative hypoactivity in the BDD group vs. controls in the lingual gyrus, precuneus, and para-hippocampal gyrus [16]. Moreover, the use of FSH mapping provides additional information about utilization of structural connections that may explain observed fMRI differences; an inference from this is that individuals with BDD may have abnormally low information transfer between both primary and secondary visual cortical regions, and within higher-order temporal lobe visual processing systems. Results must be considered in light of multiple simplifying assumptions made to aid in this initial model development, and in light of small sample size. Future models that account for feedback, inhibitory interactions, and deactivation are necessary for more accurate estimations. Nevertheless, FSH may serve as a useful tool for integrating structural connectivity with fMRI data for improved understanding of functional activation patterns.
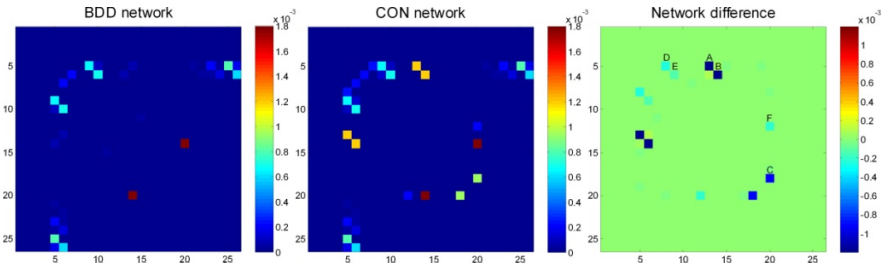

Fig. 3. Utilization matrices (left panel for BDD and middle panel for controls) and their differences (BDD – control; right panel) weighted by the mean edge strength (values in color bar). In the right panel, the most significant utilization differences (entries in blue; not utilized in BDD) represent connections between the following regions: (A) left intracalcarine cortex and left lingual gyrus, (B) right intracalcarine cortex and right lingual gyrus, (C) right temporal occipital fusiform cortex and right temporal fusiform cortex, posterior division, (D) left intracalcarine cortex and right precuneus, (E) right intracalcarine cortex and left cuneus, and (F) right para-hippocampal gyrus, posterior division and right temporal occipital fusiform cortex.
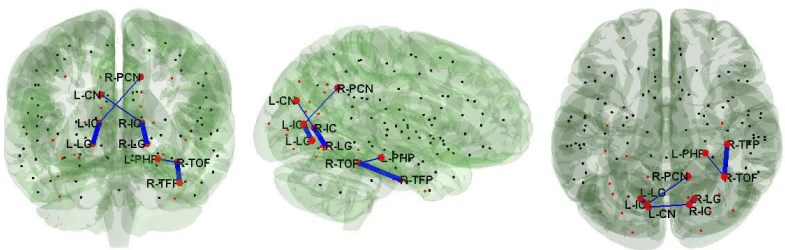

Fig. 4. Additional anatomical connections (marked in blue) utilized in the control relative to the BDD group. Here, the centers of the 26 visual system ROIs are represented by red dots. (The centers of other nodes not used in this study are shown for reference and are represented as black dots). The thicker lines correspond to connections A, B, and C, and thinner lines D, E, and F in the right panel of Figure 3. L/R indicates left/right, and IC, LG, TOF, TFP, PCN, CN, PHP denotes the intracalcarine cortex, the lingual gyrus, the temporal occipital fusiform cortex, the temporal fusiform cortex posterior division, the precuneus, the cuneus, and the parahippo-campal gyrus, respectively.

References

1. Friston, K.J., Harrison, L., Penny, W.: Dynamic causal modelling. Neuroimage 19(4), 1273–1302 (2003)
2. Saygin, Z.M., Osher, D.E., Koldewyn, K., Reynolds, G., Gabrieli, J.D., Saxe, R.R.: Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat. Neurosci. 15(2), 321–327 (2011)
3. Johansen-Berg, H., Behrens, T.E., Robson, M.D., Drobnjak, I., Rushworth, M.F., Brady, J.M., Smith, S.M., Higham, D.J., Matthews, P.M.: Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proc. Natl. Acad. Sci. USA 101, 13335–13340 (2004)
4. Passingham, R.E., Stephan, K.E., Kotter, R.: The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3(8), 606–616 (2002)
5. Chulwoo, L., Li, X., Li, K.M., Guo, L., Tianming, L.T.M.: Brain state change detection via fiber-centered functional connectivity analysis. In: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2011)
6. Skudlarski, P., Jagannathan, K., Anderson, K., Stevens, M.C., Calhoun, V.D., Skudlarska, B.A., Pearlson, G.: Brain Connectivity Is Not Only Lower but Different in Schizophrenia: A Combined Anatomical and Functional Approach. Biol. Psychiatry 68(1), 61–69 (2010)
7. Deligianni, F., Robinson, E.C., Bechmann, C.F., Sharp, D., Edwards, A.D., Rueckert, D.: Inference of functional connectivity from structural brain connectivity. In: 2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (2010)
8. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. USA 106(6), 2035–2040 (2009)
9. Varkuti, B., Cavusoglu, M., Kullik, A., Schiffler, B., Veit, R., Yilmaz, O., Rosenstiel, W., Braun, C., Uludag, K., Birbaumer, N., Sitaram, R.: Quantifying the Link between Anatomical Connectivity, Gray Matter Volume and Regional Cerebral Blood Flow: An Integrative MRI Study. PLoS One 6(4), e14801 (2011)
10. Sporns, O., Tononi, G., Edelman, G.M.: Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cereb Cortex 10, 127–141 (2000)
11. Koch, M.A., Norris, D.G., Hund-Georgiadis, M.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. Neuroimage 16, 241–250 (2002)
12. Venkataraman, A., Rathi, Y., Kubicki, M., Westin, C.F., Golland, P.: Joint Modeling of Anatomical and Functional Connectivity for Population Studies. IEEE Trans. Med. Imaging 31(2), 164–182 (2012)
13. Felleman, D.J., Van Essen, D.C.: Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cereb Cortex 1(1), 1–47 (1991)
14. Lamme, V.A., Roelfsema, P.R.: The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23(11), 571–579 (2000)
15. American Psychiatric Association: Diagnostic and statistical manual of mental disorders: DSM-IV-TR, 4th edn., vol. xxxvii, p. 943. American Psychiatric Association, Washington, DC (2000)
16. Feusner, J.D., Moody, T., Hembacher, E., Townsend, J., Mckinley, M., Moller, H., Bookheimer, S.: Abnormalities of visual processing and frontostriatal systems in body dysmorphic disorder. Arch. Gen. Psychiatry 67(2), 197–205 (2010)

17. Feusner, J.D., Townsend, J., Bystritsky, A., Bookheimer, S.: Visual information processing of faces in body dysmorphic disorder. Arch. Gen. Psychiatry 64(12), 1417–1425 (2007)

18. Feusner, J.D., Hembacher, E., Moller, H., Moddy, T.D.: Abnormalities of object visual processing in body dysmorphic disorder. Psychol. Med. 41(11), 2385–2397 (2011)

19. Iidaka, T., Yamashita, K., Kashikura, K., Yonekura, Y.: Spatial frequency of visual image modulates neural responses in the temporo-occipital lobe. An investigation with event-related fMRI. Cogn. Brain Res. 18(2), 196–204 (2004)

20. Mori, S., van Zijl, P.C.: Fiber tracking: principles and strategies - a technical review. NMR Biomed. 15(7-8), 468–480 (2002)

21. Morgan, V.L., Mishra, A., Newton, A.T., Gore, J.C., Ding, Z.H.: Integrating functional and diffusion magnetic resonance imaging for analysis of structure-function relationship in the human language network. PLoS One 4(8), e6660 (2009)

22. Jenkinson, M., Bannister, P., Brady, M., Smith, S.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17(2), 825–841 (2002)

Characterization of Task-Free/Task-Performance Brain States

Xin Zhang[1,2], Lei Guo[1], Xiang Li[2], Dajiang Zhu[2], Kaiming Li[3], Zhenqiang Sun[4],
Changfeng Jin[5], Xintao Hu[1], Junwei Han[1], Qun Zhao[6],
Lingjiang Li[5], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, China
[2] Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA
[3] Biomedical Imaging Technology Center, Emory University, Atlanta, GA
[4] The School of Electronic and Information Engineering,
Xi'an Jiaotong University, Xi'an, China
[5] The Mental Health Institute, The Second Xiangya Hospital,
Central South University, Changsha, China
[6] Department of Physics and Astronomy and Bioimaging Research Center,
The University of Georgia, Athens, GA

Abstract. Both resting state fMRI (R-fMRI) and task-based fMRI (T-fMRI) have been widely used to study the functional activities of the human brain during task-free and task-performance periods, respectively. However, due to the difficulty in strictly controlling the participating subject's mental status and their cognitive behaviors during fMRI scans, it has been very challenging to tell whether or not an R-fMRI/T-fMRI scan truly reflects the participant's functional brain states in task-free/task-performance. This paper presents a novel approach to characterizing the brain's functional status into task-free or task-performance states. The basic idea here is that the brain's functional state is represented by a whole-brain quasi-stable connectivity pattern (WQCP), and an effective sparse coding procedure was then applied to learn the atomic connectivity patterns (ACP) of both task-free and task-performance states based on training R-fMRI and T-fMRI data. Our experimental results demonstrated that the learned ACPs for R-fMRI and T-fMRI datasets are substantially different, as expected. However, a certain portion of ACPs from R-fMRI and T-fMRI datasets are overlapping, suggesting that those subjects with overlapping ACPs were not in the expected task-free/task-performance states during R-fMRI/T-fMRI scans.

Keywords: DTI, fMRI, connectivity, cortical landmarks.

1 Introduction

In the brain imaging field, resting state fMRI (R-fMRI) [1, 2] and task-based fMRI (T-fMRI) [3] have been widely employed to investigate the functional activities of the human brain in task-free and task-performance periods. However, it has been rarely

studied (as far as we know) whether or not the R-fMRI/T-fMRI data was really reflecting the subject's task-free/task-performance states, in that it is very difficult to strictly control the participating subject's mental status and their cognitive behaviors during fMRI scan sessions. For instance, if a participating subject's brain was active, e.g., in some active cognitive processes, during the R-fMRI scan, how different will this R-fMRI data be from other strict R-fMRI data acquired during task-free states? Similarly, if a participating subject's brain was in resting state, e.g. not strictly following the administered task-performance paradigm [10], how different will this T-fMRI data be from other strict T-fMRI data scanned during task-performance states? If these differences are substantial, can we quantitatively characterize and automatically differentiate those unreliable or false R-fMRI/T-fMRI data from strict R-fMRI/T-fMRI data? The answers and solutions to these questions can significantly enhance our understanding of the function mechanisms of the brain and enable us to detect and control the quality of R-fMRI/T-fMRI data in the subsequent quantitative analysis, e.g., inference of resting state networks (RSNs), functional connectivity analysis, and task-based functional region localization.

In response to the above unanswered questions, this paper presents a novel computational framework to characterize the brain's task-free and task-performance functional states by learning from both R-fMRI and T-fMRI datasets. The basic idea is that we represent the brain's functional status by whole-brain quasi-stable connectivity patterns (WQCP), and then apply a sparse coding approach to learn the atomic connectivity patterns (ACP) of both task-free and task-performance states from large-scale temporally segmented WQCPs. Notably, the integration and pooling of many WQCPs from different brains are enabled by our recently developed and validated 358 consistent cortical landmarks, or regions of interests (ROIs), in [5], which provide intrinsic structural and functional correspondences across individuals and populations. Thus, the WQCPs from different temporal segments of multiple brains can be readily pooled and effectively compared via sparse coding and representation methods, which can learn the most descriptive atomic patterns in forming a meaningful dictionary to represent and discriminate those WQCPs. Our experimental results demonstrated that the learned ACPs for R-fMRI and T-fMRI datasets are substantially different, as expected, but the overlapping ACPs suggest that certain subjects were not in the expected task-free/task-performance states and should be considered as outliers in the following steps of data analysis.

2   Materials and Methods

2.1   Overview

The flowchart of the proposed computational framework is summarized in Fig. 1. First, 358 cortical ROIs discovered and validated in our recent study in [5] are located in the brain using DTI data (green bubbles in the left panel of Fig. 1). Then, both resting state fMRI (R-fMRI) and visual-task fMRI (T-fMRI) time series for each ROI are extracted. By using a sliding time window, the dynamic functional connectivity time series between each pair of ROIs are measured and the cumulative connectivity
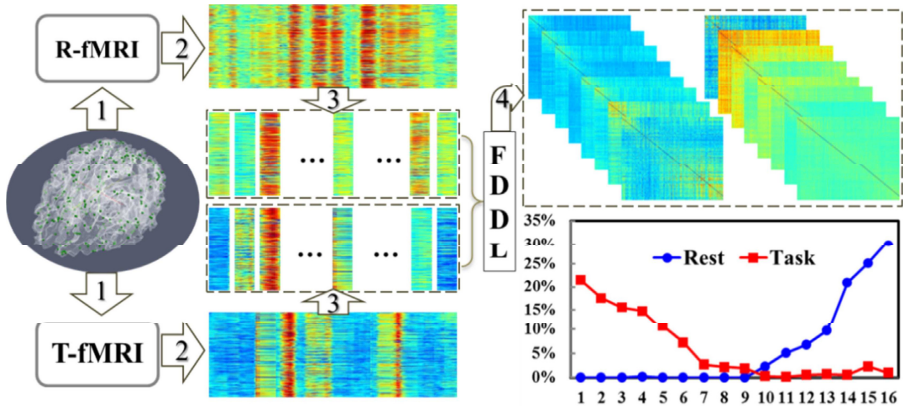
Fig. 1. The flowchart of our computational framework. (1) fMRI signal extraction for each ROI; (2) Measurement of dynamic functional connectivity strength; (3) Manual segmentation and collection of WQCP training samples; (4) FDDL sparse learning and classification.

strength of each ROI at each time point is summed. It is observed that the functional connectivity strengths are relatively stable in a continuous time period, and then the dynamic functional connectivity time series are manually segmented into quasi-stable time periods (called WQCP above), which form a set of WQCP training samples. Finally, the WQCP samples from both R-fMRI and T-fMRI datasets were combined together for sparse representation learning via the Fisher discriminative dictionary learning (FDDL) method [4].

2.2   Data Acquisition and Pre-processing

Twenty-six healthy adolescent volunteers participated in this study under IRB approvals. Multimodal DTI and fMRI datasets were acquired on a 3T GE MRI scanner. Both resting state fMRI and block-based visual task fMRI scans were acquired for each volunteer. Acquisition parameters for the scans were as follows. fMRI: 64×64 matrix, 4mm slice thickness, 220mm FOV, 30 slices, TR=2s; Visual task design and imaging parameters are referred to our recent publication [7]. DTI: 256×256 matrix, 3mm slice thickness, 240mm FOV, 50 slices, 15 DWI volumes, b-value=1000. The pre-processing of the DTI data included brain skull removal and motion correction. Both resting state and visual task-based fMRI datasets were pre-processed using the FSL FEAT.

2.3   WQCP Extraction

Based on the DTI data of each subject, we predicted the 358 cortical landmarks via the functional ROI prediction approaches in [5]. In brief, these 358 cortical landmarks were optimized to possess consistent group-wise structural connection patterns, and thus have structural and functional correspondences across individuals and populations. The left panel of Fig.1 shows an example of the distributions of the 358 cortical landmarks on a cortical surface. In particular, these 358 cortical landmarks
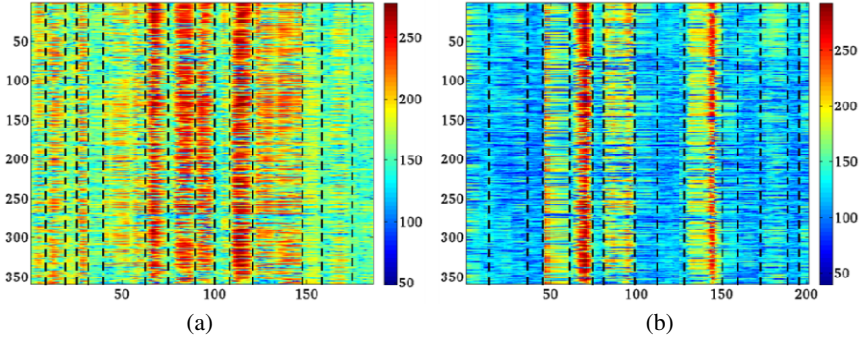
Fig. 2. The dynamic functional connectivity strength time series and WQCP segments. The horizontal axis represents time points and the vertical axis represents the cumulative functional connectivity strength of each ROI. The WQCPs are marked by the dash black lines. (a) A resting state fMRI case; (b) A visual task fMRI case.

are reproducible and predictable across different brains and most of them have been functionally annotated into meaningful networks via either benchmark task-based fMRI data or functional meta-analysis [5, 6]. Thus, this set of 358 cortical landmarks offers a common and individualized brain reference system for functional connectivity analysis, which is also adopted here in this paper.

After all the 358 cortical landmarks were located in the subject's brain, the fMRI time series for each landmark can be extracted from R-fMRI and T-fMRI datasets (Section 2.2). To investigate the temporally dynamics of the large-scale functional brain connectivities, a sliding time window approach was applied here. At each time point t, the functional connectivity between each pair of ROIs (i and j) is defined by:

$$FC_{i,j} = abs\left(Pearson\ Correlation(W_i, W_j)\right),\ FC_{i,j} = 0, \text{if } i = j; \qquad (1)$$

where $W_i = [s_{i,t}, s_{i,t+1}, \dots, s_{i,t+l-1}]$ and $W_j = [s_{j,t}, s_{j,t+1}, \dots, s_{j,t+l-1}]$. l is the length of the time window; $s_{i,t}$ and $s_{j,t}$ are the fMRI signal strengths of ROI i and ROI j at the time point t. FC is measured using the absolute value of Pearson correlation coefficient between the two l-time-points length fMRI time series W_i and W_j. After the dynamic functional connectivity time series between each pair of ROIs are obtained, the cumulative connectivity strength of each ROI is measured by summing all the functional connectivities between this ROI and all of the other ROIs. This converts the functional connectivity matrix into a representative connectivity vector at each time point. Then, we obtained a 2D dynamic functional connectivity strength matrix for each participant (see Fig. 2). In this figure, the horizontal axis represents time points and the vertical axis is the cumulative functional connectivity strength of each ROI, which is color-coded by the color-bar on the right. From Fig. 2, it can be observed that the connectivity strength keeps relatively stable in a continuous time period. Therefore, the dynamic connectivity matrix is manually segmented into a series of shorter time segments, called whole-brain quasi-stable connectivity patterns (WQCP). Then, since the functional connectivity pattern within each WQCP is quasi-

stable, each WQCP segment is averaged among the time axis, resulting in a single WQCP vector. Two experts performed segmentation work and independently checked. Finally we obtained a collection of 1149 consistent WQCP vectors from the datasets in Section 2.2. Specifically, there are 474 WQCP vectors for resting state data and 675 WQCP vectors for visual task data. All these WQCP segments and vectors were pooled together as training samples and represented using the following methods.

2.4   FDDL for Sparse Representation of WQCP

Sparse representation has been widely demonstrated to exhibit very good performance in a variety of image analysis applications such as image classification [4, 9]. Typically, there are two steps in the sparse representation based image classification method: coding and classification. In sparse representation, learning the descriptive and representative dictionary is the key. Sparse dictionary learning has been used in the brain activity and function study [11]. This paper adopted the recently developed Fisher discriminative dictionary learning (FDDL) based sparse representation methodology [4] and tailored it for our functional brain state learning applications.

Briefly, the FDDL method employs a Fisher discrimination criterion to learn a structured dictionary, based on which the classification is performed. Here, the learned dictionary is denoted by $D = [D_1, D_2, ..., D_c]$, where D_i is the sub-dictionary corresponding to the class i, and c is the total number of classes learned. Also, $A = [A_1, A_2, ..., A_c]$ represents the training WQCP vector samples, where A_i is the sub-set of the training WQCP vector samples belonging to the class i. In addition, $X = [X_1, X_2, ..., X_c]$ represents the coding coefficient matrix of A over D. The FDDL model is represented as follows:

$$J_{(D,X)} = argmin_{(D,X)}\{r(A, D, X) + \lambda_1\|X\|_1 + \lambda_2 f(X)\} \tag{2}$$

where the first term on the right $r(A, D, X)$ is called the discriminative fidelity term; the second term $\|X\|_1$ is the sparsity constraint; and the last term $f(X)$ is a Fisher discrimination constraint imposed on the coefficient matrix. λ_1 and λ_2 are scalar parameters for trade-off between sparsity and discrimination capability. Here, λ_1=0.005 and λ_2=0.05.

Specifically, there are two classifiers that can be used: global classifier (GC) and local classifier (LC) [4]. This study adopted the GC to perform the sparse coding learning and classification. For one input WQCP vector sample y, first, the sparse coding coefficients can be obtained by solving the following:

$$\hat{\alpha} = argmin_\alpha\{\|y - D\alpha\|_2^2 + \gamma\|\alpha\|_1\} \tag{3}$$

where $\hat{\alpha} = [\hat{\alpha}_1; \hat{\alpha}_2; ...; \hat{\alpha}_c]$ and $\hat{\alpha}_i$ is the coefficient vector linked to the D_i. γ is a constant parameter. Then, the sample y is attributed to the class, associated with which the sub-dictionary has the minimum representation error defined by Eq. 4:

$$e_i = \|y - D_i\hat{\alpha}_i\|_2^2 + w \cdot \|\hat{\alpha} - m_i\|_2^2 \tag{4}$$

where the first term is the reconstruction error using class i, and the second term is the distance between $\hat{\alpha}$ and m_i. m_i is the learned mean vector and w is a constant.

This study used and tailored the above FDDL methodology to train the pooled resting state and task-based WQCP vector samples to achieve a compact and meaningful dictionary. In particular, each sub-dictionary or class is associated with an atomic connectivity pattern (ACP), and here the total number of ACPs (and thus sub-dictionaries) is set to be 16, which is optimally selected based on many experiments. Afterward, we used the GC to classify the WQCP samples into 16 ACP patterns and obtained the distributions of WQCP vectors for both resting state and task samples in each ACP. Finally, each ACP pattern is described and represented by the averaged center of all of the WQCP segments and WQCP vectors belonging to this pattern.

3   Experimental Results

3.1   Distributions of 16 ACPs in R-fMRI and T-fMRI Datasets

Using methods in Sections 2.4, we obtained a combined dictionary containing 16 sub-dictionaries or ACPs. Then, each WQCP sample is classified into one ACP in correspondence with one sub-dictionary. The distributions in the 16 ACPs for both task-free and task-performance WQCP samples are drawn in Fig. 3. The horizontal axis indexes the 16 ACPs discriminated by the FDDL dictionary. The vertical axis represents the percentages of WQCP samples distributed in each ACP for both task-free and task-performance datasets.

From Fig. 3, it can be observed that the task-free and task-performance periods show quite different distribution patterns. In total, there are 94.2% of task-performance WQCP samples distributed in ACPs #1~#9, which can be considered as the task-performance ACP patterns. Also, there are 97.5% task-free WQCP samples distributed in the ACPs #11~#16, which should be considered as task-free ACP patterns. However, for ACP #10 (highlighted by

Fig. 3. The distributions of resting state and task-based WQCP samples in 16 ACP patterns

black arrow in Fig. 3), the task-free and task-performance WQCP samples overlap and the percentages for both are lower than 3%. Importantly, it is exactly these overlapped WQCP samples that imposed difficulty to decide which functional state (task-free or task-performance) it really belongs to, and we consider them as an uncertain pattern or even outliers in the other following steps of data analysis, such as identification of resting state networks (RSNs) and task-based functional region localization.
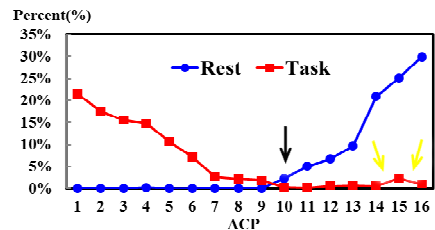
3.2   Visualization and Quantitative Analysis of ACPs

To explore more neuroscience meanings of the ACP patterns, we visualized these 16 ACPs using their averaged WQCP centers. In Fig. 4, each ACP pattern is represented by a 358×358 functional connectivity matrix corresponding to one WQCP center.
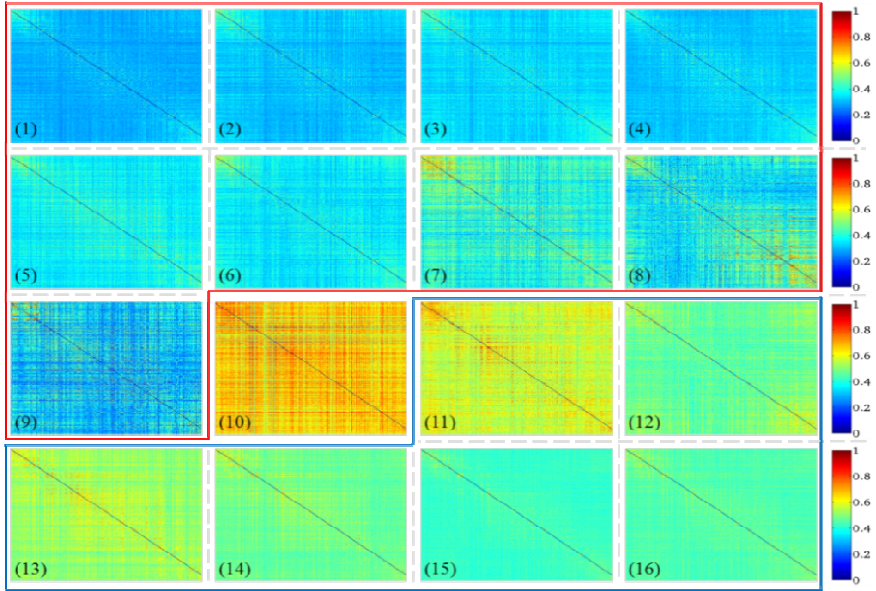

Fig. 4. Sixteen ACP patterns. It is a compact representation of task-free and task-performance states together.

In Fig. 4, ACPs #1~#9 (Fig. 3) are considered as task-performance ACP patterns (in the red frame in Fig. 4) and ACPs #11~#16 are considered to be task-free ACP patterns (in the blue frame in Fig. 4). It can be easily appreciated that ACP patterns in resting state and task-performance are quite different.
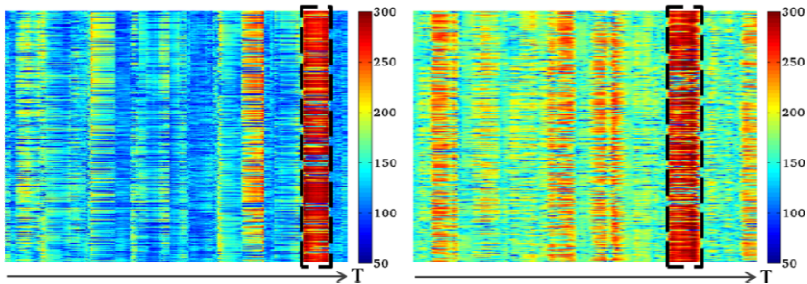

Fig. 5. WQCP samples with ACP pattern #10 in task-performance (left) and resting state (right). The ACP pattern #10 is highlighted by the dashed black box.

In total, we found 11 WQCP samples in 8 task-free subjects and 2 WQCP samples in 2 task-based subjects that were classified into the uncertain ACP #10. This pattern (Fig. 4(10)) exhibits quite high global functional connectivity, and Fig. 5 shows two examples from both T-fMRI and R-fMRI WQCP samples. In addition to the shared ACP #10 in Figs.3-4, we further investigated the potential outliers in both resting state and task-performance WQCP samples. For instance, there is one WQCP sample in a task-free subject, but it was clustered into the ACP #4, which is considered as one task-performance ACP. Importantly, we found 37 WQCP segments (out of totally 675 task-performance WQCP samples) in 17 subjects that were clustered into the task-free ACPs, as shown by the red boxes in the right side of Fig. 3 (highlighted by the yellow arrows). The quantitative

Table 1. The numbers and percentages of subjects with detected outlier ACP patterns.

ACP	Rest Num/Percent	Task Num/Percent
Pattern #1	0	-
Pattern #2	0	-
Pattern #3	0	-
Pattern #4	1/3.8%	-
Pattern #5	0	-
Pattern #6	0	-
Pattern #7	0	-
Pattern #8	0	-
Pattern #9	0	-
Pattern #10	8/30.8%	2/7.7%
Pattern# 11	-	1/3.8%
Pattern #12	-	4/15.4%
Pattern #13	-	4/15.4%
Pattern #14	-	3/11.5%
Pattern #15	-	11/42.3%
Pattern #16	-	3/11.5%

summaries are provided in Table 1. These results imply that the participants in our experiments exhibited relatively good resting performance for high quality R-fMRI data, but they did not perform equally well in visual task experiments, as 17 of them exhibited resting state ACP patterns during the task-performance scans, suggesting these subjects were not well following the administered tasks in certain periods. Thus, we should take additional caution when analyzing the task-based fMRI datasets of these 17 subjects.

4   Discussion and Conclusion

This paper presents a novel framework for quantitative characterization of task-free and task-performance functional brain states via sparse representation of whole-brain quasi-stable connectivity patterns (WQCP). Experimental results have demonstrated that though the learned ACPs for R-fMRI and T-fMRI datasets are substantially different, a certain portion of overlapping ACPs between the two datasets suggests that some subjects were not in the expected task-free/task-performance states during R-fMRI/T-fMRI scan sessions. This result has important implications in detecting and controlling R-fMRI/T-fMRI data quality for other data analysis tasks. In the future, we will examine the detailed functional connectivity patterns in all ACPs. For instance, the ACP #16 in Fig. 4 can be clustered into several functional sub-networks (Fig. 6), and it turns out that the widely replicated default mode network [1] is within

one clustered sub-network, as highlighted by the red lines in Fig. 6. This result suggests that we can possibly define and cluster resting state networks, e.g., the ones within the black boxes in Fig. 6, within each temporally quasi-stable ACP pattern, in which the temporal patterns of functional connectivities are much more homogeneous and stable than those in traditional RSN identification methods that consider the entire R-fMRI scan period [2].
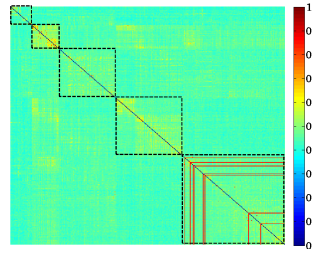

Fig. 6. The default mode network (DMN) in ACP pattern #16

References

1. Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. Proc. Natl. Acad. Sci. USA 98(2), 676–682 (2001)
2. Fox, M.D., Raichle, M.E.: Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8, 700–711 (2007)
3. Heeger, D.J., Ress, D.: What does fMRI tell us about neuronal activity? Nat. Rev. Neurosci. 3(2), 142–152 (2002)
4. Yang, M., Zhang, L., Feng, X., Zhang, D.: Fisher Discrimination Dictionary Learning for Sparse Representation. In: ICCV, Barcelona, pp. 543–550 (2011)
5. Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C., Jin, C., Wee, C.Y., Yuan, Y., Lv, P., Yin, Y., Hu, X., Duan, L., Hu., X., Han, J., Wang, L., Shen, D., Miler, L.S., Li, L., Liu, T.: DICCCOL: Dense Individualized and Common Connectivity-Based Cortical Landmarks. Cerebral Cortex (accepted, 2012)
6. Laird, A.R., Eickhoff, S.B., Kurth, F., Fox, P.M., Uecker, A.M., Turner, J.A., Robinson, J.L., Lancaster, J.L., Fox, P.T.: ALE meta-analysis workflows via the BrainMap database: Progress towards a probabilistic functional brain atlas. Front Neuroinformatics 3(23) (2009)
7. Zhang, T., Guo, L., Li, K., Jin, C., Hu, X., Cui, G., Li, L., Liu, T., Li, L., Liu, T.: Predicting functional cortical ROIs via DTI-derived fiber shape models, Cerebral Cortex (2011) (in press)
8. FMRIB Software Library, http://www.fmrib.ox.ac.uk/fsl/index.html
9. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in Neural Informattion Processing Systems (NIPS), Vancouver, vol. 19, pp. 801–808 (2007)
10. Epstein, J.N., Casey, B.J., Tonev, S.T., Davidson, M., Reiss, A.L., Garrett, A., Hinshaw, S.P., Greenhill, L.L., Vitolo, A., Kotler, L.A., Jarrett, M.A., Spicer, J.: Assessment and prevention of head motion during imaging of patients with attention deficit hyperactivity disorder. Psychiatry Res. 155(1), 75–82 (2007)
11. Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B.: Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. Inf. Process Med. Imaging 22, 562–573 (2011)

Quantitative Evaluation of Statistical Inference in Resting State Functional MRI

Xue Yang[1], Hakmook Kang[2], Allen Newton[3], and Bennett A. Landman[1,3]

[1]Electrical Engineering, Vanderbilt University, Nashville, TN, USA, 37235
[2]Biostatistics, Vanderbilt University, Nashville, TN, USA, 37235
[3]Institute of Image Science, Vanderbilt University, Nashville, TN, USA, 37235

Abstract. Modern statistical inference techniques may be able to improve the sensitivity and specificity of resting state functional MRI (rs-fMRI) connectivity analysis through more realistic characterization of distributional assumptions. In simulation, the advantages of such modern methods are readily demonstrable. However quantitative *empirical* validation remains elusive *in vivo* as the true connectivity patterns are unknown and noise/artifact distributions are challenging to characterize with high fidelity. Recent innovations in capturing finite sample behavior of asymptotically consistent estimators (i.e., SIMulation and EXtrapolation - SIMEX) have enabled direct estimation of bias given single datasets. Herein, we leverage the theoretical core of SIMEX to study the properties of inference methods in the face of diminishing data (in contrast to increasing noise). The stability of inference methods with respect to *synthetic* loss of *empirical* data (defined as *resilience*) is used to quantify the empirical performance of one inference method relative to another. We illustrate this new approach in a comparison of ordinary and robust inference methods with rs-fMRI.

Keywords: fMRI connectivity analysis, validation, resampling, resilience.

1 Introduction

When the brain is at rest (i.e., not task driven), functional networks produce correlated low frequency patterns of activity that can be observed with resting state fMRI (rs-fMRI). These correlations define one measure of *functional connectivity* which may be estimated by voxel-wise regression of activity in a seed region against that of the remainder of the brain [1]. The sensitivity and specificity of connectivity inference techniques hinge upon valid models of the *noise* in the observed data. Structured violations of the noise models due to local flow, bulk motion, distortions, or other artifacts can invalidate the methods traditionally used for inference [2]. Modern robust and non-parametric methods remain valid over a broader range of disturbances, but come with the cost of reduced power when the traditional methods would be appropriate. Therefore, a quantitative approach for comparing inference methods (and preprocessing pipelines leading to inference) is an essential analytical tool.

Several approaches for evaluating fMRI inference methods have been proposed. When repeated datasets are available, one can measure the reproducibility of

estimated quantities when inference is applied to each dataset separately [3]. In task-based fMRI, cross-validation resampling procedures have been used to assess spatial patterns of reproducibility and temporal predictability for fMRI of task activities with the held-back samples [4]. More recent approaches for defining inference performance have considered the inference procedure as a classifier between the patterns of task activity and image intensity [5]. Yet, these advanced approaches are not applicable to rs-fMRI, and to date, no methods have been proposed to quantify relative performance of rs-fMRI inference methods based on typically acquired datasets (i.e., without large numbers of repeated scans for a single subject).

Herein, we propose a new inference comparison approach based on the *resilience* of the inference estimator. We apply this new technique to characterize ordinary and robust inference of rs-fMRI data. This approach does not require acquisition of additional data and is suitable for evaluation on isolated datasets as well as groups.

2   Theory

SIMEX is a statistical method that can be adapted to create resilience measures for inference in rs-fMRI. The principle behind SIMEX is that the expected value of an estimator diverges smoothly with increasing noise levels, therefore, the mean degree of corruption can be estimated by extrapolating a trend of divergence when synthetic noise is added to *empirical* data [6]. In our context, it is not reasonable to add noise because the noise distributions are uncertain — especially in the context of outliers. If we apply the SIMEX assumption of smooth convergence in this case, we can probe the marginal reduction in sensitivity of an estimator by removing data.

We define *resilience* as the ability of an inference method to maintain a consistent connectivity estimate despite a reduction in data. Over the time course of an rs-fMRI experiment (5-10 mins), the active brain regions vary. Hence, reproducibility of inferences based on sampled time periods is not meaningful. Therefore, we focus on decimating the sampling rate (**Fig 1**). The resilience of t-value estimates is quantified by two summary metrics: (i) the average absolute value of change in t-value with decimation level (i.e., slope), (ii) the average variance of the estimated metrics. The slope of t-value is computed by averaging the individual slopes between decimation levels.

2.1   Regression Models

rs-fMRI data can be analyzed with a first order autoregressive model, AR(1), for a weakly stationary time series [7],

$$y_i = X\beta_i + e_i, \ \ e_i \sim N(0, \sigma_i^2 V) \tag{1}$$

where y_i is a vector of intensity at voxel i, X is the design matrix, β_i is a vector of regression parameters at voxel i, and e_i is a non-spherical error vector. The correlation matrix V is estimated using Restricted Maximum Likelihood (ReML) and β is estimated on the whitened data (i.e., the "OLS" approach). Alternatively, a robust estimator (e.g., the "Huber" M-estimator [8]) may be applied after whitening.
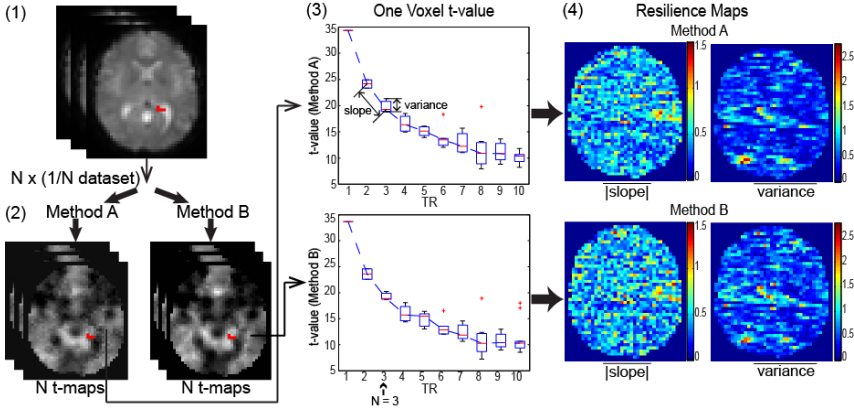
Fig. 1. Resilience features capture the stability of an inference method to data decimation. An rs-fMRI dataset (1) is temporally decimated into subsets; each inference method (2) is applied independently to each subset; voxel-wise statistics (3) are estimated; and the parameter maps (4) capture spatial dependencies.

Both the OLS and Huber methods are available within the SPM software [9]. Herein, we used the Huber method with the tuning constant chosen for 95% asymptotic efficiency when the distribution of observation error is Gaussian distribution [10].

3   Methods and Results

3.1   One Voxel Simulation

Resilience aims to capture the performance of statistical inference methods on empirical data where the true correlations are unknown. If the true correlations are known, one could directly calculate the type I and type II errors. To explore how our definition of resilience relates to the type I error and the type II error, we performed single voxel simulation using an AR(1) model,

$$y = \beta_0 + \beta_1 x + e, \quad e \sim N(0, V) \tag{2}$$

where x was simulated as region of interest (ROI) time course containing 200 time points with an approximately uniform distribution between 10 and 20 (arbitrary units), e was distributed as zero mean autoregressive Gaussian noise with standard deviation equaling 10% of the mean y value and normalized correlation 0.2. The null hypothesis $H_0: \beta_1 = 0$ was tested using a t-test. In separate simulations, β_1 was assigned to 0 for specificity exploration and 0.8 for sensitivity exploration. The type I and type II errors are calculated based on p-value ($p < 0.05$). To simulate structured outliers, Rician distributed noise ($Rice(0,25)$) was added to the y values corresponding to x values with the lowest or largest intensity. Rician noise is used to reflect the distribution of noise in MR magnitude images. The number of outliers was swept between 0 and 10 using 10^3 Monte Carlo repetitions each. The covariance matrix was

assumed to be known, and OLS and Huber inference were performed independently after whitening. Resilience metrics were calculated with three data partitions (degraded up to ¼ of the dataset). Fig. 2 compares type I error, type II error, and resilience as a function of number of outliers; significant differences were evaluated using the Wilcoxon signed-rank test.

Both OLS and Huber controlled the type I and type II errors when there were no outliers. For the resilience, the mean absolute slope and the mean variance of OLS and Huber are not significantly different when $\beta_1 = 0$ without outliers. When $\beta_1 = 0.8$ without outliers, the mean absolute slope from OLS is larger than Huber while the mean variance is smaller. These results are in agreement with the known behavior that robust methods are not as powerful as OLS when assumptions are met.

When considering outliers, Huber resulted in lower mean absolute slope and mean variance than OLS (for $\beta = 0$). When $\beta = 0.8$, the mean absolute slope of Huber was was higher (due to higher t-statistics with all data), but Huber yield lower variance estimates. Hence, we must consider both the mean absolute slope and mean variance in consideration of estimator performance as these are complementary measures. The mean variance from OLS increases when outliers appear because some decimation samples include outliers while others do not. In contrast, Huber is more resistant to outliers so that the variances are relatively constant. The resilience results show less significance (e.g., last columns in the mean variance in Fig. 2) which is concordant with the decrease in the proportion of the differences in absolute errors. In summary, the resilience is strongly correlated with the type I and type II errors.

3.2   Empirical 3T rs-fMRI Experiment

Eleven rs-fMRI of healthy subjects were acquired at 3T using EPI (197 vol, FOV = 192 mm, flip θ = 90°, TR/TE = 2000/25 ms, 3x3x3 mm, 64x64x39 voxels) [11]. Prior to analysis, all images were corrected for slice timing artifacts and motion artifacts using SPM8 (University College London, UK). All time courses were low pass filtered at 0.1 Hz using a Chebychev Type II filter, spatially normalized to Talairach space, spatially smoothed with an 8 mm FWHM Gaussian kernel, linearly detrended, and de-meaned. Two voxels inside the right primary motor cortex for each subject were manually selected as the ROI by experienced researchers through exploring the unsmoothed images and comparing with the standard atlas. The design matrix for the general linear model was defined as the ROI time courses, the six estimated motion parameters, and one intercept. To create whole-brain connectivity maps, every labeled brain voxel underwent linear regression using the design matrix followed by a one sided t-test on the coefficient for the ROI time courses.

For each subject, the whole dataset (197 scans) was subsampled. First, the TR value was set to be 4 s (TR = 2 s in the original dataset), the 197 time series fMRI scans were divided into two subsamples, one containing 99 scans and the other containing 98 scans. Similarly, the TR value was set to be 6 s to obtain three subsamples. This procedure was repeated with a TR value of 8 s. Thus, we have one original dataset and three collections of subsampled datasets for each subject. The resting state fMRI analysis was performed on each dataset in SPM8 using OLS and Huber inference.
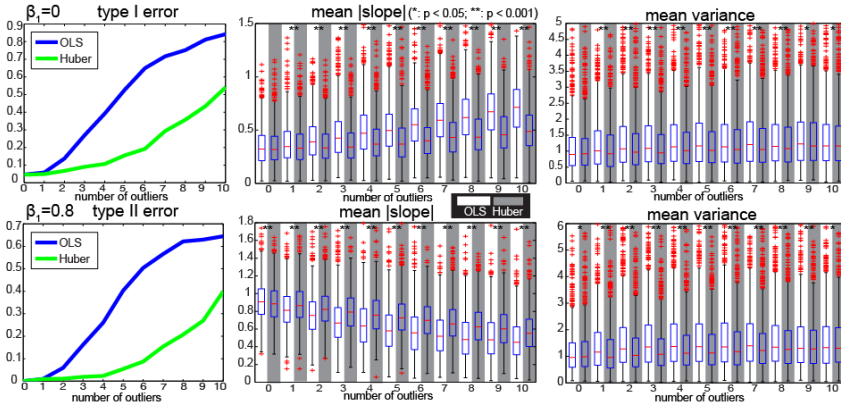
Fig. 2. One voxel simulation. Y axes are indicated by panel titles. The first row shows the results when $\beta_1 = 0$ and the second row shows the results when $\beta_1 = 0.8$. When the number of outliers increases, the type I error and the type II error of OLS increases more rapidly than Huber M-estimator. The boxplots in the white background display the results from OLS and the boxplots in the gray background are the results from Huber M-estimator.

To quantitatively compare the resilience of these two methods, the mean absolute slope and the mean variance are evaluated across the gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). To evaluate all subjects, we calculated the mean of the mean absolute slope value and the mean of the mean variance in each brain region for each subject. The significance of differences between the ordinary and the robust estimation method were tested using the Wilcoxon signed-rank test.

The mean absolute slope and the mean variance from OLS are smaller than those of Huber (Fig. 3). Over the 11 subjects, the mean absolute slope of OLS is not significantly different while the mean variance is significantly smaller. Thus, the resilience metrics confirm expectations that OLS is a superior inference technique for high quality empirical data (i.e., when distributional assumptions are appropriate).

3.3   Empirical 3T rs-fMRI Experiments with Outliers

To illustrate the use of resilience in the presence of outliers, a dataset with outliers was simulated by increasing the WM intensity of the empirical 3T rs-fMRI dataset described in section **3.2**. The simulation results show that three outliers are enough to tell the difference between OLS and Huber estimation so we selected three scans with low ROI intensity and added random positive noisy images inside the WM region to simulate outlier scans. Noisy Rician images are created with σ at 10% of the mean intensity and spatially smoothed at 8 mm FWHM Gaussian kernel. One slice of an outlier image is displayed in Fig. 4 (compare with Fig. 3).

We applied the same connectivity analysis method (OLS and Huber) and the same resilience calculation method (TR from 2s to 8s, mean absolute slope and mean variance in GM, WM and CSF) described in section 3.2 (Fig. 4).

The connectivity maps illustrate that the OLS method lost substantial power in the vicinity of the seed voxels when outliers were introduced whereas the Huber method preserved detection. In terms of resilience, the Huber approach resulted in significantly smaller mean absolute slope and mean variance in the WM. We also noted smaller mean absolute slope and smaller mean variance from the robust method than the ordinary method in GM regions. The relative improvement of the Huber performance in GM may due to the spatially pooled covariance estimation. In CSF, the mean absolute slope from OLS is larger while the mean variance is smaller for the example subject and across subjects. Noting that there are no outliers in the CSF it is reasonable that the performance of Huber is not better than OLS. The resilience results here suggest that the robust estimation method outperforms the OLS method if outliers are present.
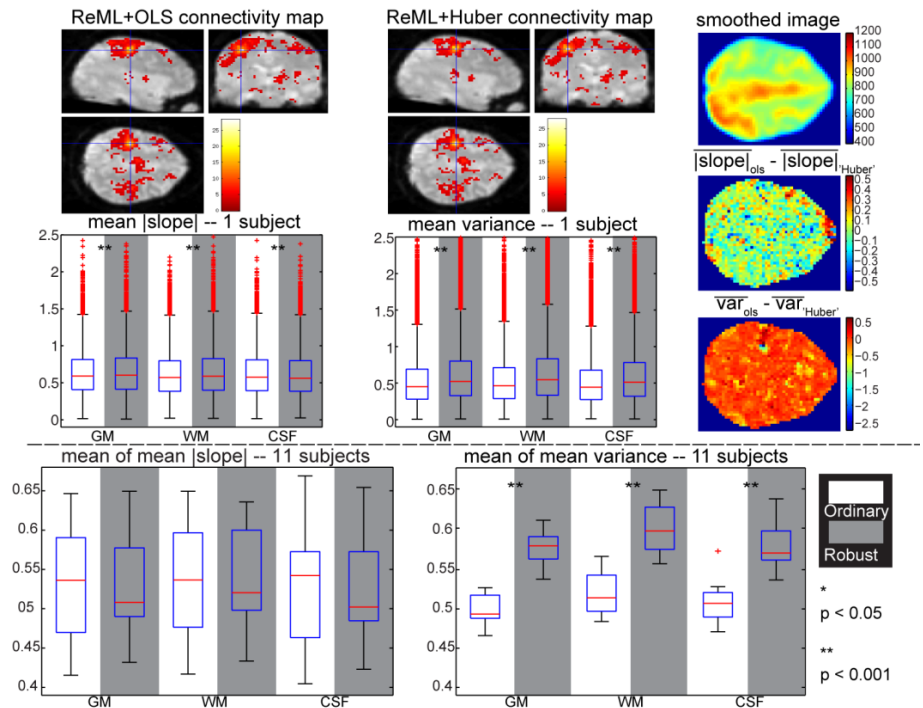

Fig. 3. Resilience results for empirical 3T rs-fMRI analysis. The upper plots present results for a representative subject and the lower plots display the results across the 11 subjects. The first row shows the connectivity maps estimated by the OLS and Huber methods ($p < 0.001$, 5 voxels extent threshold to exclude noise). The blue crosshairs indicates one voxel inside the ROI. The right column displays one slice of the smoothed image from one scan (top) the difference of the mean absolute slope (middle), and the difference of the mean variance (bottom) for the same slice. The mean absolute slope and the mean variance from OLS (white background) and Huber (gray background) across GM, WM and CSF regions are shown in the second row. In the second half, the mean of the mean absolute slope and the mean of the mean variance across eleven subjects are displayed. Significant differences based on the Wilcoxon signed-rank test are indicated by the asterisks.

4   Discussion

The proposed resilience metrics provide a quantitative basis on which to compare inference methods. The simulation results suggest that a comparison of methods based on resilience would yield similar conclusions as one based on the type I and type II errors. It is reassuring to see that resilience also indicates that OLS would outperform a Huber inference approach when the data quality is high (as in the publicly available dataset under study), whereas a Huber approach would outperform OLS in cases when outliers are present. As rs-fMRI is applied to ever more challenging anatomical targets (i.e., those requiring high spatial and temporal resolution and/or using ultra-high field imaging), the achievable signal to noise ratio decreases and the propensity for artifacts increases [12]. Hence, it is becoming increasingly more important to quantitatively determine which inference methods are appropriate.
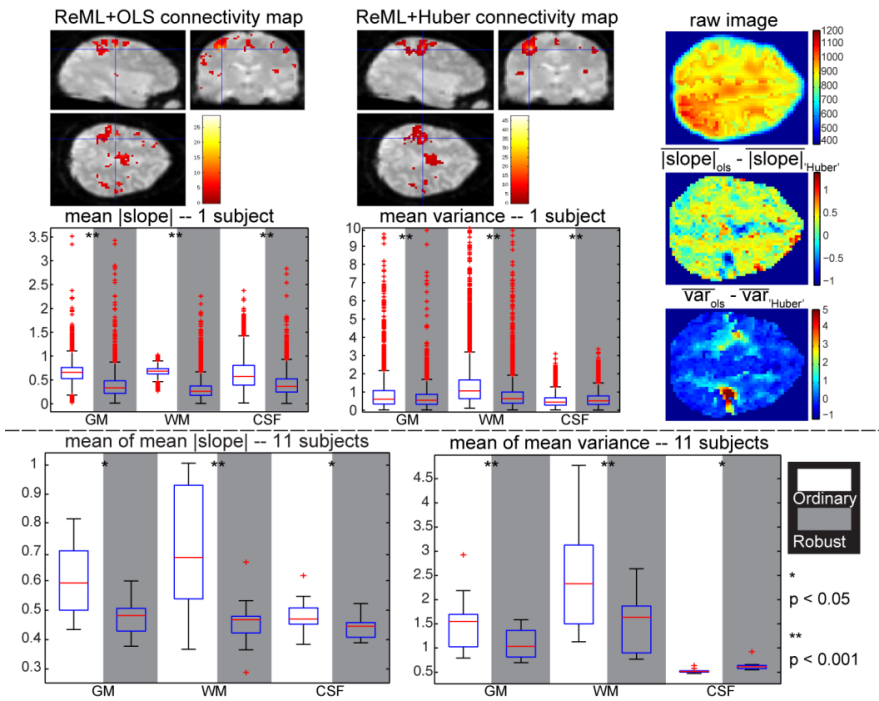

Fig. 4. Resilience results for 3T resting state fMRI with simulated outliers. The upper plots present the results for the same subject shown in **Fig. 3** and the lower plots display the results across the 11 subjects. The first row shows the connectivity maps estimated by the OLS and Huber methods (p < 0.001, 5 voxels extent threshold to exclude noise). The right column displays (top) one outlier image from one scan for the same slice shown in **Fig. 3**, (top) the difference of the mean absolute slope and (bottom) the difference of the mean variance. The mean absolute slope and the mean variance from the ordinary and the robust method across GM, WM and CSF regions are shown in the second row. Below, the mean of the mean absolute slope and the mean of the mean variance across eleven subjects are displayed. Significant differences calculated with the Wilcoxon signed-rank test are indicated by asterisks.

In summary, we have presented a novel approach for quantifying inference methods based on empirical data. Herein, we evaluated the resilience of the ordinary (OLS) and a robust method (Huber) for both simulated and empirical data. Resilience provides a simple, but powerful method for comparing a proxy for accuracy of inference approaches in empirical data where the underlying true value is unknown. Continued exploration of metrics based on resilience criteria promises to provide a fruitful avenue for comparative characterization of inference stability and "quality."

Note that if two inference methods yield different t-values when all data are considered, the one that has a higher starting t-value will have a higher mean absolute slope even if both methods degrade at the same rate. Hence, in the regions of true association (i.e., $\beta \neq 0$), the variance measure is likely of greater interest as it reflects degraded inference consistency. Yet, in regions that lack an association (i.e., $\beta = 0$), the slope measure would reflect on anonymous changes in t-value which could be attributed to "non-robust" influences. Consideration of data-adaptive combinations of these metrics would be area of fruitful investigation.

References

1. van den Heuvel, M.P., Hulshoff Pol, H.E.: Exploring the brain network: a review on resting-state fMRI functional connectivity. European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology 20, 519–534 (2010)
2. Penny, W.D., Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, New York (2006)
3. Genovese, C.R., Noll, D.C., Eddy, W.F.: Estimating test-retest reliability in functional MR imaging I: Statistical methodology. Magnetic Resonance in Medicine 38, 497–507 (1997)
4. Strother, S.C., Anderson, J., Hansen, L.K., Kjems, U., Kustra, R., Sidtis, J., Frutiger, S., Muley, S., LaConte, S., Rottenberg, D.: The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. Neuroimage 15, 747–771 (2002)
5. Afshin-Pour, B., Soltanian-Zadeh, H., Hossein-Zadeh, G.A., Grady, C.L., Strother, S.C.: A mutual information-based metric for evaluation of fMRI data-processing approaches. Human Brain Mapping 32, 699–715 (2011)
6. Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.: Measurement error in nonlinear models: a modern perspective. CRC Press (2006)
7. Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J.: Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465–483 (2002)
8. Huber, P.J., Ronchetti, E., MyiLibrary: Robust statistics. Wiley Online Library (1981)
9. Yang, X., Beason-Held, L., Resnick, S.M., Landman, B.A.: Biological parametric mapping with robust and non-parametric statistics. NeuroImage 57, 423–430 (2011)
10. Holland, P.W., Welsch, R.E.: Robust regression using iteratively reweighted least-squares. Communications in Statistics-Theory and Methods 6, 813–827 (1977)
11. Shehzad, Z., Kelly, A., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B.: The resting brain: unconstrained yet reliable. Cerebral cortex 19, 2209 (2009)
12. Hutton, C., Josephs, O., Stadler, J., Featherstone, E., Reid, A., Speck, O., Bernarding, J., Weiskopf, N.: The impact of physiological noise correction on fMRI at 7T. Neuroimage (2011)

Identifying Sub-Populations
via Unsupervised Cluster Analysis
on Multi-Edge Similarity Graphs

Madhura Ingalhalikar[1], Alex R. Smith[1], Luke Bloy[1], Ruben Gur[2],
Timothy P.L. Roberts[3,*], and Ragini Verma[1]

[1] Section of Biomedical Image Analysis,
University of Pennsylvania, Philadelphia, PA, USA
{Madhura.Ingalhalikar,Ragini.Verma}@uphs.upenn.edu
[2] Brain Behavior Laboratory, University of Pennsylvania, Philadelphia, PA, USA
[3] Lurie Family Foundation's MEG Imaging Center, Department of Radiology,
Children's Hospital of Philadelphia, Philadelphia, PA, USA

Abstract. Pathologies like autism and schizophrenia are a broad set
of disorders with multiple etiologies in the same diagnostic category.
This paper presents a method for unsupervised cluster analysis using
multi-edge similarity graphs that combine information from different
modalities. The method alleviates the issues with traditional supervised
classification methods that use diagnostic labels and are therefore unable
to exploit or elucidate the underlying heterogeneity of the dataset under
analysis. The framework introduced in this paper has the ability to em-
ploy diverse features that define different aspects of pathology obtained
from different modalities to create a multi-edged graph on which cluster-
ing is performed. The weights on the multiple edges are optimized using a
novel concept of 'holding power' that describes the certainty with which
a subject belongs to a cluster. We apply the technique to two separate
clinical populations of autism spectrum disorder (ASD) and schizophre-
nia (SCZ), where the multi-edged graph for each population is created
by combining information from structural networks and cognitive scores.
For the ASD-control population the method clusters the data into two
classes and the SCZ-control population is clustered into four. The two
classes in ASD agree with underlying diagnostic labels with 92% accuracy
and the SCZ clustering agrees with 78% accuracy, indicating a greater
heterogeneity in the SCZ population.

1 Introduction

Classifying subjects based on their underlying pathology, brain structure, behav-
ior and cognition is an important step towards creating biomarkers. However,
pathologies like ASD and other neuropsychiatric disorders are defined over a
spectrum and the severity of the disease may vary within a population thus

making the data highly heterogeneous. Different modalities, like imaging, neurocognitive scores etc., may characterize different aspects of this heterogeneity to different degrees. This paper presents a method for unsupervised cluster analysis of populations using multi-edge similarity graphs that combine information of population heterogeneity from different modalities, producing classes that are more representative of population variability.

Traditional superivised classification methods, utilize predefined diagnostic labels for the subjects for training [1], [2], and hence new subjects can only be classified into one of these diagnostic categories, thereby overlooking the underlying heterogeneity of the pathology. These also require a large sample size to capture all the variability.

Unsupervised classification or clustering are powerful techniques for self-organized categorization of the underlying data [3] without the use of diagnostic labels. Earlier studies have used various clustering algorithms in classifying tissue types or segmenting lesions in brain images [4]. However, with diseases now being grouped into a variety of classes, population analysis using such unsupervised methods is gaining interest in the neuroimaging community [5]. Recently, one study by Filipovych et al. performed semi-supervised clustering on datasets [6]. The method was limited to use information from single imaging modality and thus overlooked other components of the pathology.

Ideally, for precisely grouping a subject into a certain category, information from diverse imaging modalities, psychological scores, demographics as well as genetic information can be combined together and clustered without utilizing clinical diagnostic information. Such a technique will thus provide a comprehensive grouping and aid in understanding the underlying patterns of pathology.

With such an aim, in this paper we present a novel method that employs unsupervised clustering on multiple features to better understand the underlying data structure and to identify coherent subpopulations if any. We define each subject by its structural networks computed from DTI data and a battery of cognitive scores. The nodes of the structural network are clustered and similarity between subjects is computed using the variance of information metric between these clusters. For the cognitive battery, the similarity is computed using Euclidean distance. Thus we get a dual-edged similarity graph, in which subjects represent the nodes and the similarities represent the edges. We then perform unsupervised spectral clustering on the linear combination of these similarity graphs. The optimal linear combination is defined via the concept of 'holding power' that provides a basis of certainty with which a subject belongs to a particular cluster. The weights on the individual similarity graph quantify the participation of that feature in the clustering process. We apply this method to two datasets, with autism and schizophrenia pathologies, respectively, to determine the ability of our method in identifying homogeneous subpopulations.

2   Methods

Here we describe the method of clustering on multi-edge graphs created on the population with the edges defining inter-subject similarity based on different
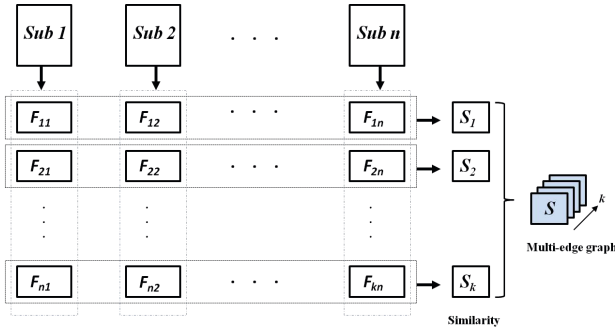
Fig. 1. This diagram depicts the technique used to create a multi-edge graph. Diverse features $(F_1, ..., F_k)$ are extracted for each subject that may include information from connectivity matrices, image intensity values, genomic data or cognitive scores. The similarity matrices $(S_1, .., S_k)$ are then computed over all the subjects for each feature. Together these matrices form the multi-edge graph which defines various facets of similarity (edges) between the nodes (subjects).

modalities such as imaging and cognitive scores and define the concept of holding power of each node in the multi-edge graph, which represents the power with which the subject (represented by a node) belongs to a cluster.

2.1 Unsupervised Clustering on Multi-edge Graphs

Consider a similarity graph $S = (V, E)$, created over a population where the subjects are represented by the nodes V and the similarities based on the modalities are defined by the edges E. Each edge e in the graph represents the connection (similarity in our case) between two nodes with a weight w ($w \in R$). For a multi-edge graph, with k number of linkages (each representing a modality) the edges between any two nodes can be described by $w \in R^k$.

The goal of unsupervised clustering on such a graph is to partition the graph, utilizing information from all possible linkages/modalities, such that nodes with tighter connections (with high similarity) cluster together while nodes with loose connections (low similarity) are placed in different clusters.

Clustering a multi-edge graph is a challenging problem as each edge type conveys different information and thus can cause the subjects to cluster differently. We address this challenge by flattening the graph into a single edged graph by employing a linear function where $f = \sum \alpha_i S_i$ where $i = (1, 2, ..k)$ and α_i is the weight on all the edges of S_i. Our focus here is to obtain good clustering by using information from all the available features. Therefore the problem translates to finding values of α that optimize the clustering quality. To quantify the quality of clustering, we use the concept of pull and holding power as proposed by Rocklin et al. [7].

We begin with flattening the multi-edge graph where the α_i are chosen randomly under the condition $\sum \alpha_i = 1$. Unsupervised clustering is then applied

to this linear combination $(\alpha_1 S_1 + \alpha_2 S_2 +, .., \alpha_k S_k)$. Any type of unsupervised clustering method can be applied (e.g. affinity clustering, graph clustering etc). Since our goal here is to demonstrate the importance of multi-edge graph, we use a standard spectral clustering algorithm [8] that clusters the nodes in M clusters. For each node $v \in V$, the 'pull' to each cluster C_m in $C = (C_1, C_2, \ldots, C_M)$ is then defined as the average weights of edges between node v and the nodes categorized in cluster C_m. Therefore, for a given set of coefficients α, the pull on node v is defined by equation 1 where x is the number of nodes categorized in cluster C_m.

$$P_\alpha(v, C_m) = \frac{1}{x} \sum_{w=(u,v)\in E, u \in C_m} w(\alpha) \qquad (1)$$

The holding power $H_\alpha(v)$ for each node, is then defined as the pull of the cluster to which the node belongs minus the next largest pull among other clusters. Thus, for a node v in cluster C_m, the holding power is defined by equation 2.

$$H_\alpha(v) = P(v, C_m) - \max_{C_i \in C, C_i \neq C_m} P(v, C_i) \qquad (2)$$

If the holding power is positive, then we can say that the node is held in the right cluster. Thus, to achieve superior clustering that justifies the position of each node in that cluster, we can maximize the holding power as well as the number of nodes with positive holding power. For easier implementation of holding power in optimization routines, the function can be smoothed by using $\mathrm{atan}(H_\alpha(v))$.

2.2 Multi-edge Graphs from Structural Networks and Cognitive Scores

In our study the two edge weights that we use are computed from the full brain structural connectivity networks the cognitive scores of each subject.

The structural network is a $n*n$ connectivity matrix where n regions of interest (ROI's) in gray matter are defined and the edge computation between two nodes is based on the density of white matter fibers between these nodes. Here, we explain the similarity between the structural networks of two subjects via the similarity between the community structures of the structural networks for the two subjects. Obtaining communities in the structural network is essentially equivalent to performing clustering on the nodes of the structural network. Here, we use a standard unsupervised spectral clustering algorithm for obtaining the community structure for each subject [8]. We then compute the distance between two subjects or two clusterings by using a variation of information (VI) metric [9] which is based on the mutual information between two clustering's [9].

Consider $C_i = (C_i^1, C_i^2, ..., C_i^K)$ to be a clustering on subject i and $C_j = (C_j^1, C_j^2, ..., C_j^{K`})$ to be the clustering on subject j and let n be the total number of nodes. Then $P(C, k) = \frac{|C^k|}{n}$ is the probability that the node is in cluster C^k in a clustering C_i and in clustering C_j is $P(C_i, C_j, k, l) = \frac{|C_i^k \cap C_j^l|}{n}$. The entropy of

information in C_i is defined by equation 3, while the mutual information shared by C_i and C_j is given by equation 4.

$$H(C_i) = -\sum_{k=1}^{K} P(C_i, k) \log P(C_i, k) \tag{3}$$

$$I(C_i, C_j) = \sum_{k=1}^{K} \sum_{l=1}^{K'} P(C_i, C_j, k, l) \log P(C_i, C_j, k, l) \tag{4}$$

The variation of information metric that describes the distance between two clusterings is then defined by equation 5.

$$d_{VI}(C_i, C_j) = H(C_i) + H(C_j) - 2I(C_i, C_j) \tag{5}$$

The similarity between the battery of cognitive scores creates the second edge of the multi-edge graph. The edge ij of the similarity matrix is the Euclidean distance between the cognitive scores of subject i and subject j. For a vector of length r the Euclidean distance is computed by equation 6.

$$d(cog_i, cog_j) = \sqrt{\sum_{l=1}^{r} (cog_{il} - cog_{jl})^2} \tag{6}$$

To convert the variation of information distance matrix as well as the distance matrix of cognitive test scores, to similarity matrices, we use the negative exponential of the distance as proposed by Shepard [10]. Thus, we now have a multi-edge similarity graph on which unsupervised spectral clustering can be performed. For optimizing the holding power, optimizers such as gradient descent can be applied. However, in our case, since it's only a dual-edged graph, a simple grid search performs reasonably. We then apply the weights computed at the maximum holding power to obtain the final clustering on the population.

The clustering is then validated against the ground truth diagnosis. The validation enables us to understand the performance of the technique as well as provides insight into the heterogeneity of the dataset.

3   Results

3.1   Simulated Data

We consider a population of 8 subjects, described by 3 modalities. We simulated three $8*8$ similarity matrices (S_1, S_2, S_3) each representing the similarities described by a specific modality. As can be seen from Fig 2, S_1 and S_3 were designed to impose a clustering (with modality S_3 characterizing the first 5 subjects better and modality S_1 characterizing the last 3 subjects better), while S_2 was diffuse. We then combined these matrices to create a 3 edged graph (as described in section 2.1). The weights (α) on these similarity matrices were optimized using a grid search method, to maximize the total holding power and
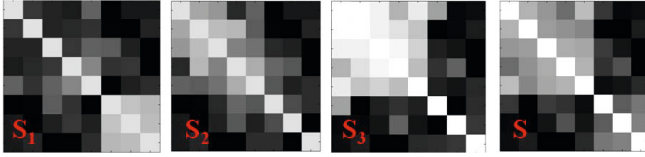
Fig. 2. Figure shows the simulated similarity matrices (S_1, S_2 and S_3) which form the multi-edge graph. S is the linear combination of S_1, S_2 and S_3 with maximum holding power and the 2 clusters in S are evident.

minimize the number of nodes (subjects) with negative holding power, that is, nodes that are mis-clustered. The maximum holding power was achieved at 0.4 for S_1 and 0.6 for S_3, while the weight on S_2 was zero since it did not add anything to maximize the holding power, thereby establishing the feasibility of the holding power concept. The optimization implicitly puts more weight on the matrices with stronger connections, thus choosing S_1 and S_3 and removing S_2. Spectral clustering on the final similarity matrix S caused it to cluster into 2 classes with the first five nodes in one cluster and last three in other cluster which matches the underlying clustering of the data.

3.2   Real Data

Two separate datasets were used in the unsupervised clustering:

- The SCZ dataset consisted of 29 female controls (CNT) and 23 female age matched patients with schizophrenia. The DWI images were acquired on Siemens 3T scanner with b=1000 s/mm^2 and 64 gradient directions. Neurocognitive testing was carried out on all the subjects and the speed and accuracy of memory, emotion, reasoning, and executive functioning were recorded.
- The ASD dataset consisted of 33 participants with ASD and 21 age matched typically developing controls (TD's). The DWI images were acquired on Siemens 3T scanner with b=1000 s/mm^2 and 30 gradient directions. The cognitive and psychological tests included verbal IQ, Social Responsiveness Scale (SRS), Social communication questionnaire (SCQ), Clinical evaluation of language fundamentals (CELF), Full scale IQ and Autism diagnostic observation schedule (ADOS) and perceptual reasoning index (PRI).

Computing the Structural Network. Cortical parcellation and sub-cortical segmentation of all the subjects was obtained using Freesurfer [11] on structural T1 images, and a total of 78 ROI's were extracted to represent the nodes of the structural network. These labels were then transferred to the diffusion space via intrasubject affine transformation. Probabilistic fiber tracking [12] was employed to determine the percentage of streamlines that exit ROI i and enter ROI j. The conditional probability is given by $p_{ij} = \frac{S_{i \to j}}{S_i}$, where $S_{i \to j}$ denotes the number of fibers reaching j, and S_i is the number of streamlines seeded in i.

We normalize p_{ij} by the active surface area R_i of the ROI i to get connectivity measure P_{ij}, which accounts for different sizes of the active seed region. This measure quantifies connectivity such that $P_{ij} \approx P_{ji}$ which upon averaging, gives an undirected weighted connectivity matrix.

We then perform clustering on the computed structural networks. We apply spectral clustering to each subject and then compute the VI matrix (Section 2.2, equation 5) for the two datasets. The similarity matrix from the cognitive scores is computed using the Euclidean distance as described in Section 2.2. Figure 3 shows the similarity matrices for the two datasets. The red lines divide the matrix to reveal patients and controls. The cognitive score of the ASD produces a 'visual clustering' in the matrix (d), however the psychological testing scores for SCZ (b), do not produce such a distinctive difference as is evident in the color coding of the matrix. The structural connectivity similarity matrices, do not impose such acute clustering.

Unsupervised Clustering on Multi-edge Graph. We performed spectral clustering on the multi-edge graphs for both the datasets, such that the holding power is maximized with minimum number of subjects having negative holding power. The weights at maximum holding power for SCZ data were 0.55 for structural network and 0.45 for the psychological scores, suggesting that the combination of information aided in the clustering process. The entire dataset was clustered into 4 groups with two clusters representing the SCZ patients and 2 clusters representing the controls with 78% accuracy.

For ASD data, the maximum holding power was achieved with a weight of 0.2 for structural and 0.8 for psychological scores. The spectral clustering on this linear combination split the data into two clusters: one with ASD and other with TD with 92% accuracy. This suggested that although DTI did not add much, the combination of information was important to maximize the hold of each subject in that cluster. When only psychological scores were used, the data was split into 3 clusters, where the third cluster consisted 4 subjects (2 ASD and 2 TD).
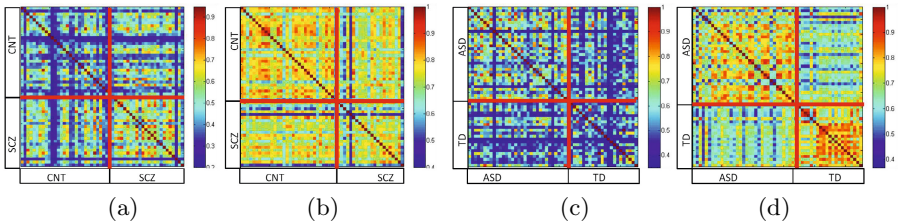

Fig. 3. (a) Similarity matrix computed from structural connectivity network for SCZ (b) similarity matrix computed from cognitive scores in SCZ (c) Similarity matrix computed from structural networks in ASD (d) Similarity matrix computed from cognitive scores in ASD. The horizontal and vertical red lines show the control-patient division in the group.

4 Conclusion

In this paper, we have created a novel technique for unsupervised clustering that creates population based multi-edged graphs using different modalities and features. Spectral clustering on these graphs in conjunction with maximizing the holding power of each subject in a cluster was used to identify population subgroups. The method was validated on simulated data and then applied to datasets with ASD and schizophrenia. We found two inherent clusters in the ASD data while schizophrenia data was more heterogeneous with four inherent clusters. A direct interpretation of such clusters is non-trivial, but it is our working hypothesis that relevant features found using this methodology will map onto the clinical space of cognitive scores. In future, we plan on expanding this idea for defining the heterogeneity index over a patient population in a spectrum disorder like ASD.

References

1. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: classification of morphological patterns using adaptive regional elements. IEEE Trans. Med. Imaging 26(1), 93–105 (2007)
2. Ingalhalikar, M., Parker, D., Bloy, L., Roberts, T.P.L., Verma, R.: Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of asd. Neuroimage 57(3), 918–927 (2011)
3. Duda, R., Hart, P., Stork, D.: Pattern Classification. Wiley, New York (2001)
4. Xu, R., Wunsch, D.C.: Clustering algorithms in biomedical research: a review. IEEE Rev. Biomed. Eng. 3, 120–154 (2010)
5. Sabuncu, M., Balci, S., Golland, P.: Discovering Modes of an Image Population through Mixture Modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 381–389. Springer, Heidelberg (2008)
6. Filipovych, R., Resnick, S.M., Davatzikos, C.: Semi-supervised cluster analysis of imaging data. Neuroimage 54(3), 2185–2197 (2011)
7. Rocklin, M., Pinar, A.: Latent Clustering on Graphs with Multiple Edge Types. In: Frieze, A., Horn, P., Prałat, P. (eds.) WAW 2011. LNCS, vol. 6732, pp. 38–49. Springer, Heidelberg (2011)
8. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. U S A 103(23), 8577–8582 (2006)
9. Meilă, M.: Comparing Clusterings by the Variation of Information. In: Schölkopf, B., Warmuth, M.K. (eds.) COLT/Kernel 2003. LNCS (LNAI), vol. 2777, pp. 173–187. Springer, Heidelberg (2003)
10. Shepard, R.N.: Toward a universal law of generalization for psychological science. Science 237(4820), 1317–1323 (1987)
11. Fischl, B., et al.: Automatically parcellating the human cerebral cortex. Cereb Cortex 14(1), 11–22 (2004)
12. Behrens, T.E.J., et al.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. 6(7), 750–757 (2003)

Geodesic Information Flows

M. Jorge Cardoso[1], Robin Wolz[2], Marc Modat[1], Nick C. Fox[3],
Daniel Rueckert[2], and Sebastien Ourselin[1,3]

[1] Centre for Medical Image Computing (CMIC), University College London, UK
[2] Visual Information Processing Group, Imperial College London, UK
[3] Dementia Research Centre (DRC), University College London, UK

Abstract. Homogenising the availability of manually generated information in large databases has been a key challenge of medical imaging for many years. Due to the time consuming nature of manually segmenting, parcellating and localising landmarks in medical images, these sources of information tend to be scarce and limited to small, and sometimes morphologically similar, subsets of data. In this work we explore a new framework where these sources of information can be propagated to morphologically dissimilar images by diffusing and mapping the information through intermediate steps. The spatially variant data embedding uses the local morphology and intensity similarity between images to diffuse the information only between locally similar images. This framework can thus be used to propagate any information from any group of subject to every other subject in a database with great accuracy. Comparison to state-of-the-art propagation methods showed highly statistically significant ($p < 10^{-4}$) improvements in accuracy when propagating both structural parcelations and brain segmentations geodesically.

1 Introduction

Since the advent of open imaging databases, researchers have struggled with the fact that extra sources of clinical, structural and anatomical information are only available on a small subset of the data. These sources of information (e.g. manual parcelations, anatomical landmarks, tissue priors, pathological classification) are usually scarce since they require large amounts of human interaction. Ideally, one would like to be able to automatically extrapolate and propagate this information to morphologically dissimilar datasets in a coherent manner.

In neuroimage analysis, the best example of information propagation is multi-atlas segmentation. Many researchers have shown that propagating structural parcelations from multiple sources, by mapping them to new unseen data using image registration and then fusing the candidate parcelations, provides a good estimation of the true underlying parcelation [1,2]. However, in the case of limited and morphologically clustered source of information, like the 30 young control subjects with an associated parcelation of 83 key brain areas provided by Hammers et al. [3], structural parcelation propagation can be problematic. As these parcelations are defined only on young controls with normal anatomy, it is non trivial to directly map this information to morphologically dissimilar

and pathological subjects [4,5] without introducing large errors. More recently, Wolz *et al.* [6] introduced the LEAP approach (learning embeddings for atlas propagation) for brain segmentation. In LEAP, a low dimensional representation of the data is used to find morphologically similar datasets. This morphological similarity is then used to gradually diffuse the segmentation of the brain from the 30 Hammers atlases [3] to pathological Alzheimer's diseased patients via morphologically similar intermediate datasets, greatly increasing the segmentation accuracy. However, as the region-of-interest size increases, the morphological embedding becomes less localised, potentially resulting in a decrease in performance. A similar framework, but for geodesic image registration, was also introduced by Hamm *et al.* [7] with the GRAM (geodesic registration on anatomical manifolds) method. This family of step-wise propagation algorithms will become increasingly relevant with the availability of larger and larger databases. Ideally, one would like to slowly diffuse any information from its source to all the other images in a database in an unbiased manner.

This work presents an algorithm where information is propagated along the geodesic path of the local data embeddings. These local embeddings are represented as a voxel-wise implicit undirected graph with a heat kernel based information reconstruction weight that uses both the local image similarity and local image morphology as a metric. The proposed algorithm not only allows information (structural parcelations, tissue atlases, *etc.*) to be transferred between anatomically disparate images with great accuracy, but also allows the estimation of geodesic distance maps that represent the local degree of confidence and extrapolation of the propagated information. Overall, as the reconstruction weights are both symmetric and inverse consistent, the framework can be used to propagate any information from any subject (or group of subjects) to every other subject in a dataset using an unbiased information flow.

2   Methods

This section will first introduce the mathematical framework and the undirected graph for geodesic information flow, followed by the introduction of the distance metric between images. Finally, the geodesic information propagation step is presented for two types of information: numerical and label attributes.

2.1   The Implicit Local Data Embeding

Let a set Y with R images be the full set of observed *grey-valued* anatomical data with the a-th image of this set denoted by Y_a. Each image Y_a is going to be a vector of size N_a, with its i-th voxel denoted by $Y_{a,i}$.

In order to embed the observed data within a manifold, one normally starts by finding a distance between each pair of images. This distance provides insights about the global similarity between the images, and subsequently, about the manifold structure of the data. Theoretically, this global embedding assumes that one can represent the space spanned by the full data in a low dimensional

space. However, due to the complexity of the data, the dimensions of the manifold can lack interpretability and usefulness. For example, Gerber *et al.*[8] explored the manifold structure of the space of brain images and concluded that the first dimension of the manifold represents global ventricular expansion due to disease and ageing, while the second dimension meaning is described as "less obvious". Ideally, one would like to be able to capture the local manifold structure of the brain on a spatially constrained neighbourhood and not the global brain morphology. With one manifold representation per voxel, one would be able to describe the local brain morphology and similarity as a measure of distance between two mapped locations, and respective neighbourhoods, in two different images. However, due to computational and memory requirements, this problem is untractable. As an example, only to store a pairwise distance matrix for one single image at every voxel and assuming a set of 120 neighbouring images with average size 200^3, one would need approximately 400GB of memory. Furthermore, the memory requirements will grow proportionally to R^2, where R is the number of datasets. Thus, one cannot have an explicit representation of the manifold at the voxel level. In this work, instead of constructing an explicit representation of the manifold, we implicitly represent the manifold through the local neighbourhood graph of each data point. This greatly reduces both computational complexity and memory requirements, making the problem tractable and linearly scalable with the number of datasets R in the database.

Let D be a set of distance matrices, with the a-th matrix of this set denoted by D_a. Here, D_a will be an $N_a \times (R-1)$ matrix describing the distance between the image a and each one of the remaining $(R-1)$ images at every sample position i. More specifically, $D_{a \to b}(i)$ will contain the distance between the i-th sample of image Y_a and its corresponding sample in image Y_b. We now introduce a threshold d_t over these distances. By doing so, one can now build an undirected graph where the neighbourhood of each data-point is restricted only to the data-points with $D_{a \to b}(i) < d_t$. Note that the graph is undirected only if the distances are a semi-metric (subadditivity is not required).

In this work, one does not need to explicitly represent the full graph. In order to solve the information diffusion problem at a given location, one only needs to keep track of the graph neighbourhood at that specific location, visually shown in Fig. 1 - right. Here, the realm of observations (the blue connections) from the data point in bold is limited by its direct neighbouring nodes with distances below d_t. Under this undirected graph assumption, a heat kernel decay function $W_{a \to b}(i)$ is then used to diffuse the information [9]. This kernel is defined as

$$W_{a \to b}(i) = \begin{cases} e^{-\frac{D_{a \to b}(i)}{t}} & D_{a \to b}(i) < d_t \\ 0 & else \end{cases} \tag{1}$$

with t being a heat kernel temperature that will determine the speed and the distance the information can diffuse. In this work, we set $d_t = t$, meaning that the choice of t will determine both the maximum distance traveled by the information and the amount of information diffusion that occurs at each iteration.
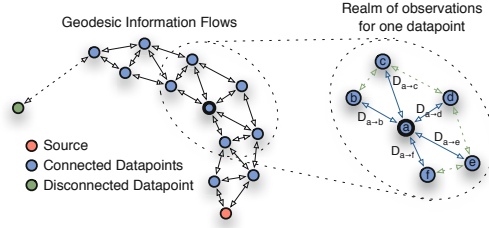
Fig. 1. Left) Implicit manifold with the neighbourhood defined as all the data points within a certain distance. Note that if the manifold is sparse, some data points might be disconnected. Right) Diagram representing the observed and unobserved connections (in blue and green respectively) and distances from the standpoint of the data point a.

2.2 The Distance Metric

The heat kernel decay function is based on the assumption that one can calculate a distance between two nodes in the graph. This distance should be at least a semi-metric, respecting both the coincidence and separation axioms and symmetry. In a medical imaging framework the local distance between images should take into account both local morphology and local image similarity. Gerber et al. [8] proposed to use coordinate transformations as a distance metric. These coordinate transformations map an image Y_a to an image Y_b by finding the optimal transformation $T_{a \to b}$ that maximises some cost function. In order to be a semi-metric, this coordinate transformation has to be symmetric, inverse consistent and diffeomorphic. In our work, we use a symmetric variant of a non-rigid free-form registration algorithm as described in [10]. Under the symmetry and diffeomorphism constraints, the transformation $T_{a \to b} = T_{b \to a}^{-1}$ and $T_{a \to b} \circ T_{b \to a} = \mathrm{Id}$, with T^{-1} being the inverse of the transformation, $\circ$ being the composition operator and Id the identity transformation. In order to remove the smoothly varying local affine component of the transformation that characterises the global anatomical shape differences, the low frequency component of the transformation is removed. From the high-frequency version of the transformation, one can then find the displacement field $F_{a \to b}$ that describes how much a point i in Y_a had to move to match the corresponding point j in Y_b.

Even though this displacement field will describe the morphological differences between the brains, we also combine it with an intensity similarity metric in order to assess the local similarity between the images after transformation [11]. This similarity term is necessary to characterise both the local differences in tissue appearance due to pathology (e.g. damaged WM in dementia) and also some possible local registration errors. The local similarity between an image Y_a and an image Y_b transformed by $T_{b \to a}$, denoted by $L_{a \to b}$, can be calculated as the local sum of squared differences (LSSD) between the intensity in these images, using a cubic B-spline kernel as a local smoothing function. We combine the two metrics together by setting $D_a(i,b) = \alpha L_{a \to b}(i) + (1 - \alpha)F_{a \to b}(i)$, with α being a relative weight, meaning that both a low displacement and a low LSSD are

necessary to obtain a low distance $D_a(i, b)$ between images. In this work $\alpha = 0.5$. Optimisation of α will be part of future work.

2.3   Geodesic Information Flows

The two previous sections have defined the neighbourhood graph and the distance metric. This section will make use of the graph structure to introduce the concept of propagating information between neighbouring nodes of the graph.

Let I be a set of vectors, with the a-th vector of this set denoted by I_a. Each vector I_a has its i-th element denoted by $I_{a,i}$, representing the associated information at location i for image Y_a. Assuming that the information to be propagated is only available in a subset of images within the full database, only some of the vectors I_a will be defined. Now, let an indicator vector K_a, indexed by $K_{a,i}$, characterise the status of the information at location $I_{a,i}$. Here, $K_{a,i}$ is set to 2 if $I_{a,i}$ is a source of information, set to 1 if the information has diffused and reached the location $I_{a,i}$ or 0 if the information does not exist at location $I_{a,i}$. As the realm of observations at each spatial location is limited by its closest neighbours, the best approximation for data flow is given by a normalised weighted sum of the information available within the neighbourhood. Thus,

$$I_{a,i} = \frac{\sum_{\forall b \in K_{b,j} > 0} W_{a \to b}(j) I_{b,j}}{\sum_{\forall b \in K_{b,j} > 0} W_{a \to b}(j)} \tag{2}$$

with j being the spatially transformed coordinate i into the space of image I_b, mapped using the previously described transformation. The information flow is thus governed by the heat kernel-derived weights $W_{a \to b}(i)$. If the set $b \in (K_{b,j} > 0)$ is not empty, the information in $I_{b,j}$ can then be propagated to $I_{a,i}$. In this case, $K_{a,i}$ is set to 1, else $K_{a,i}$ is set to 0. Note that Eq. 2 is only valid for floating point data propagation like a probabilistic atlas. The same equation can be re-formulated in a weighted label fusion scheme, by making $I_{a,i}$ equal to $p(I_{a,i}, l)$, representing the probability that location i in image a has label l, and by making $I_{b,j}$ equal to $p(I_{b,j}, l)$, representing the probability that location j has label l in image b. Eq 2 is solved iteratively for all samples where $K_{a,i} < 2$, until all are marked as solved. The number of times $I_{a,i}$ could not be solved because $K_{b,j} > 0 \; \forall \; b$ is an empty set, represents the number of steps through the graph's geodesic path necessary to transport the information from its source. The number of steps represent the amount of extrapolation of information, where larger extrapolation should result in lower accuracy.

Finally, because the temperature t will determine the neighbourhood size and consequently the existence of a connection from every information source to all the targets, the geodesic information flow is solved multiple times for several values of t. This temperature t is varied between 5 and 1 with decrements of 0.5. In a similar fashion to an annealing process, the information at the lowest possible temperature is kept as the answer, as the distance traveled by each extrapolation step is the lowest.

3 Validation

The data used in this work, depicted in Fig. 2, is compromised of two sets: 30 young controls with associated structural parcelation of 83 key structures [3] (http://www.brain-development.org); 90 subjects from the ADNI database (http://adni.loni.ucla.edu), subdivided into 30 controls, 30 Mild Cognitive Impairment (MCI) and 30 Alzheimer's diseased (AD) patients with associated manual segmentations of the brain. The current work aims at homogenising databases under the assumption that extra information is only available on a subset of the data. From these sources of information, measuring the information extrapolation accuracy will always be limited by the anatomical and pathological variability within the full dataset and by the range of available segmentations. Furthermore, the most complex sources of information, like the 30 young controls with full brain parcelations, are simply not available in pathological subjects. This makes the validation anecdotal for untested morphologies. The proposed validation will thus have two components. First, the overlap accuracy of multi-label information propagation will be estimated and compared to MAPER [4] using a leave one out approach on the 30 young controls. Then, the accuracy of information extrapolation accuracy will be characterised by propagating the brain segmentations from the elderly control group to the MCI and AD patients.

Multi-label Propagation Accuracy: The accuracy of propagating informa-tion through a geodesic path was compared to MAPER [4], a direct information fusion method based on majority voting. As the amount of parcelations available for validation is limited, a leave-one-out cross validation was performed only on the 30 young controls that have manual brain parcelations. One should note that the limited availability of segmentations restricts the range of morpholog-ical variability in the propagation, thus not representing the real performance when segmenting morphologically dissimilar subjects.

In this paper, the Dice score was used as a measure of accuracy. The mean Dice scores per structure for the the leave-one-out cross validation are shown in Table 1. Out of 83 structures, 15 structures had a significantly higher Dice score using the Geodesic information Flow when compared to MAPER, while only two structures (lingual gyrus and superior parietal gyrus) where better segmented in MAPER. The mean Dice score over all structures and all patients for the
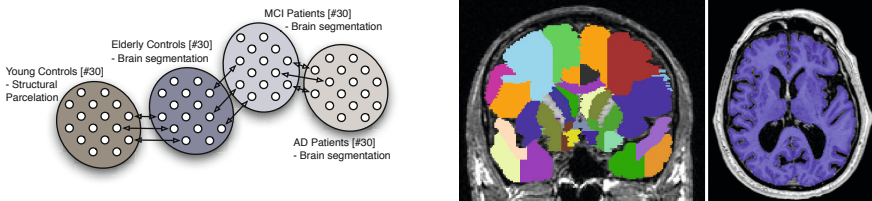

Fig. 2. Left) All the sets of data used in this work and their associated information. Right) One dataset from the Hammers atlas overlayed with its associated structural parcelation; an ADNI AD subject with the associated brain segmentation

Table 1. Mean Dice coefficient for a set of key structures, comparing the proposed method (GIF) with MAPER [4]. Statisticaly higher mean Dice is shown in bold font. Only a limited number of structures are shown due to restrictive space availability.

Structure	Unilateral Structures		
	GIF	MAPER	p-value
All Structures	0.8179	0.8089	$< 10^{-4}$
Corp. callos.	0.8802	0.8674	$< 10^{-4}$
Brainstem	0.9534	0.9377	$< 10^{-4}$
Structure	Left Side			Right Side		
	GIF	MAPER	p-value	GIF	MAPER	p-value
Hippocampus	0.8439	0.8335	0.0048	0.8251	0.8211	0.2036
Amygdala	0.8263	0.7922	$< 10^{-4}$	0.8241	0.7830	$< 10^{-4}$
Caudate nucl.	0.8983	0.8923	0.0380	0.9004	0.8955	0.0478
Nucleus acc.	0.7581	0.6834	$< 10^{-4}$	0.7317	0.6707	$< 10^{-4}$
Putamen	0.9069	0.8916	$< 10^{-4}$	0.9107	0.8959	$< 10^{-4}$
Thalamus	0.9210	0.8879	$< 10^{-4}$	0.9205	0.8852	$< 10^{-4}$
Pallidum	0.8562	0.7661	$< 10^{-4}$	0.8554	0.7672	$< 10^{-4}$

proposed method (0.8197) was significantly higher ($p < 10^{-4}$) than in MAPER (0.8089). An example of the propagation to a highly atrophied subject from the ADNI database is shown in Fig. 3 (right).

Information Extrapolation Accuracy: In the previous subsection, the accuracy of propagating information through a geodesic path was limited to a morphologically similar set of subjects. Thus, the previous validation does not capture the ability to extrapolate information to anatomically disparate subjects. The information extrapolation accuracy is thus assessed by using only a subset (the elderly control group) of all the manual brain segmentations. This morphologically clustered set of data is then used to segment both the MCI and AD groups. The proposed geodesic propagation algorithm is compared to a direct propagation algorithm based on the locally weighted majority voting algorithm [12]. The results are presented in Fig. 3 (left). The mean (std) Dice score for the proposed geodesic method was 0.940(0.009) and 0.947(0.008) for the AD an MCI groups respectively while for the direct method, the mean (std) Dice score was 0.934(0.009) and 0.942(0.008) for the AD an MCI groups respectively.
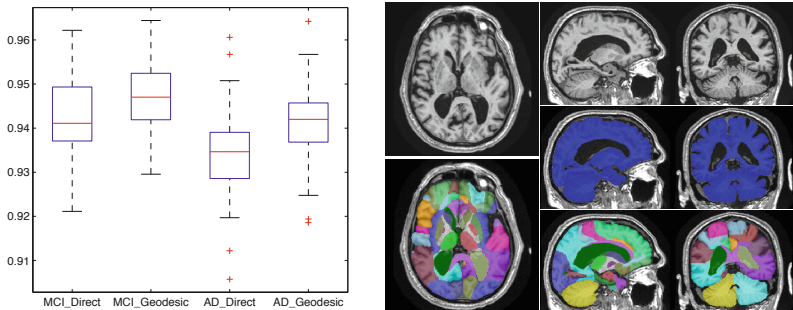

Fig. 3. Left) Dice scores for direct and geodesic propagation of brain mask. Right) An example of the propagation of both the structural parcelation and brain segmentation to an highly atrophied AD subjects (ID:1281) from the ADNI database. Note the correct ventricle segmentation and the smooth deep grey matter parcelation.

This represents a statistically significant ($p < 10^{-4}$) increase in segmentation accuracy when using a two-tailed paired t-test for statistical comparison. Note that, due to the lack of post-processing and the limited size of the training set, one should not compare these results with other brain segmentation methods.

4   Conclusion

This work presents an algorithm where information is geodesically propagated through a local implicit neighbourhood graph. Application to structural parcellation and brain segmentation propagation has demonstrated the significant ($p < 10^{-4}$) advantages of the proposed framework when compared to state of the art methods. Overall, the proposed framework can be used to propagate any information from a group of subjects to every other subject in a dataset. All the software and results used for this work is available at http://niftyseg.sf.org.

References

1. Rohlfing, T., Russakoff, D.B., Maurer, C.R.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Transactions on Medical Imaging 23(8), 983–994 (2004)
2. Aljabar, P., Heckemann, R.A., Hammers, A., Hajnal, J.V., Rueckert, D.: Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. NeuroImage 46(3), 726–738 (2009)
3. Hammers, A., Chen, C., Lemieux, L.: Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a stereotaxic space. HBM (2007)
4. Heckemann, R.A., Keihaninejad, S., Aljabar, P.: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage (51), 221–227 (2010)
5. Cardoso, M.J., Modat, M., Keihaninejad, S., Cash, D., Ourselin, S.: Multi-STEPS: Multi-label Similarity and Truth Estimation for Propagated Segmentations. In: MMBIA, pp. 153–158 (2012)
6. Wolz, R., Aljabar, P., Hajnal, J.V., Hammers, A., Rueckert, D.: Alzheimer's Disease Neuroimaging Initiative: LEAP: learning embeddings for atlas propagation. NeuroImage 49(2), 1316–1325 (2010)
7. Hamm, J., Ye, D.H., Verma, R., Davatzikos, C.: GRAM: A framework for geodesic registration on anatomical manifolds. MedIA 14(5), 633–642 (2010)
8. Gerber, S., Tasdizen, T., Joshi, S., Whitaker, R.: On the Manifold Structure of the Space of Brain Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 305–312. Springer, Heidelberg (2009)
9. Kondor, R.I., Lafferty, J.: Diffusion kernels on graphs and other discrete input spaces. In: ICML (January 2002)

10. Modat, M., Cardoso, M.J., Daga, P., Cash, D., Fox, N.C., Ourselin, S.: Inverse-Consistent Symmetric Free Form Deformation. In: Dawant, B.M., Christensen, G.E., Fitzpatrick, J.M., Rueckert, D. (eds.) WBIR 2012. LNCS, vol. 7359, pp. 79–88. Springer, Heidelberg (2012)
11. Souvenir, R., Pless, R.: Image distance functions for manifold learning. Image and Vision Computing 25(3), 365–373 (2007)
12. Yushkevich, P.A., Wang, H., Pluta, J., Das, S.R., Craige, C., Avants, B.B., Weiner, M.W., Mueller, S.: Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. NeuroImage 53(4), 1208–1224 (2010)

Group-Wise Consistent Parcellation of Gyri via Adaptive Multi-view Spectral Clustering of Fiber Shapes

Hanbo Chen[1], Xiao Cai[2], Dajiang Zhu[1], Feiping Nie[2],
Tianming Liu[1], and Heng Huang[2]

[1] Department of Computer Science and Bioimaging Research Center,
The University of Georgia, Athens, GA, USA
[2] Department of Computer Science and Engineering,
University of Texas at Arlington, Arlington, Texas, United States

Abstract. In-vivo parcellation of the cerebral cortex via non-invasive neuroimaging data has been in active research for years. A variety of model-driven and/or data-driven computational approaches have been proposed to parcellate the cortex. However, two fundamental common issues in these parcellation methodologies are the features or attributes used to define boundaries between cortical regions and the establishment of correspondences of the parcellated regions across different brains. This paper uses a novel DTI-derived fiber shape feature for the parcellation of cortical gyrus into fine-granularity segments. The gyral parcellation is formulated and solved as a surface vertex clustering problem, in which fiber shape feature similarity is used to define the distances between vertices. Then, we designed and applied a novel multi-view spectral clustering algorithm to group the vertices into group-wise consistent gyral segments across different brains. The experimental results showed that the precentral and postcentral gyrus, as two test-beds, can be consistently parcellated into 10 segments on both hemispheres across different subjects. Evaluation studies using benchmark task-based fMRI and cortical landmarks demonstrated the effectiveness and validity of the proposed methods.

Keywords: cortical parcellation, multi-view clustering, fiber shape, gyri.

1   Introduction

In-vivo parcellation of the human cerebral cortex based on neuroimaging data, e.g., MRI/DTI/fMRI, has been extensively studied because of its significant importance in basic and clinical neurosciences. In general, cortical parcellation result can be used for the definition of anatomically/connectionally/functionally meaningful structures and for measurement of their biological properties. Current approaches to cortical parcellation can be broadly categorized into two classes: model-driven and data-driven methods. In model-driven methods, atlas-based warping methods [1]-[2] are widely used to parcellate a subject's cortex by transforming an expert-labeled atlas to the subject's space. In data-driven methods, geometrical, morphological, connectional or functional features can be used to guide the parcellation procedure, e.g., in [3], a sulcal parcellation algorithm was developed based on the cortical folding patterns.

In this paper, we propose a novel approach for fine-granularity cortical gyrus parcellation via group-wise multi-view spectral clustering of fiber shape patterns. The basic idea is that each cortical gyral region is represented by its DTI-derived fiber connection shape patterns, which is able to capture the global structural connectivity pattern and thus is predictive of brain function. Then, the representative nodes of each gyrus from different brains are roughly aligned and each subject represents a view. A novel multi-view spectral clustering algorithm is designed to cluster these gryal nodes into group-wise consistent segments with correspondences across individuals. Here, the similarity matrix of the gyral nodes in each subject is modeled and represented as a separate view in the context of multi-view clustering problems. The major methodological contribution of this work is that the proposed method achieved meaningful parcellation of the gyrus and the establishment of correspondences of the parcellated segments across individuals simultaneously. The neuroscience insight gained from this work is that the seemingly very variable cortical gyrus and its fiber connections can be consistently parcellated into meaningful and common segments. In the future, the derived consistent gyral segments can be used as common ROIs (regions of interests) for a variety of neuroscience applications, such as mapping of structural and functional connectivities in healthy brains and diseased ones.

2   Methods

The computational pipeline is summarized in Fig. 1. The major novel steps are in Figs. 1c and 1d. In Fig. 1c, the gyrus is represented by discrete nodes on its extracted crest line and each gyral node is described by rich and functionally meaningful DTI-derived fiber connection patterns. In Fig. 1d, the gyral nodes from multiple brains are clustered into group-wise consistent segments by a novel multi-view spectral clustering algorithm. Finally, these DTI-derived consistent segments are validated by benchmark cortical landmarks and task-based fMRI activations on the gyri.
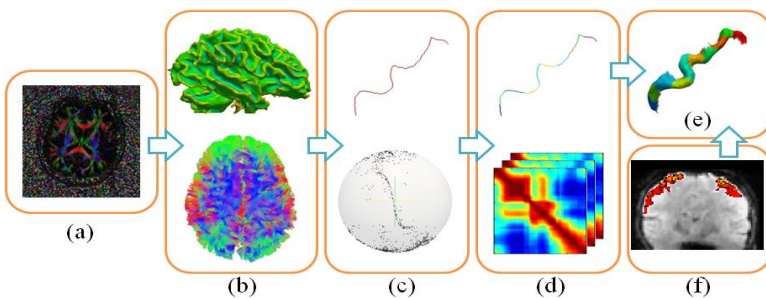

Fig. 1. The computational pipeline of group-wise multi-view clustering of fiber shape patterns for gyral parcellation. (a) DTI raw data. (b) Pre-processing including tissue segmentation, surface reconstruction, and fiber tracking. (c) Extraction of gyral crest lines and corresponding fiber shape features. (d) Group-wise clustering based on multi-view spectral clustering. (e) Gyral parcellation result. (f) Functional activation map for validation.

2.1   Data Acquisition and Preprocessing

Two separate datasets are used for this study. In the first dataset, 6 healthy volunteers were scanned with both DTI and task-based fMRI. The parameters of DTI is as follows: 128×128matrix, 2mm slice thickness, 256mm FOV, 60 slices, TR=15s, ASSET=2, 30 DWI gradient directions, 3 B0 images, b-value=1000. For fMRI scans, the motor tasks were designed using similar paradigms in [4], in which elbow, lip and ankle movements were performed respectively. The parameters are as follows: 2mm isotropic images at TR/TE=3000/25ms, FOV=256mm×256mm covering the motor cortex. 144 volumes (6 cycles) of fMRI were collected for each individual subject. In the second dataset, DTI was acquired from 10 healthy volunteers. The parameters of DTI are the same as the first dataset. The preprocessing steps are referred to [6].

2.2   Gyri Segmentation

We performed semi-automatic segmentation of two gyri, including precentral gyrus and postcentral gyrus in each hemisphere on the above reconstructed cortical surfaces from DTI. The gyral crest line (Fig. 2b) is defined as the shortest geometric path weighted by a value that is inversely proportional to the maximum principal curvature of surface (Fig. 2a) between two gyral end points on the surface. After the crest line has been selected, we resampled the crest line with 100 nodes with constant distance (Fig. 2c). After resampling, each crest line is composed by a sequence of 100 gyral nodes. After this, a rough correspondence is established between gyri across subjects by their node index, as illustrated in Fig. 2d.
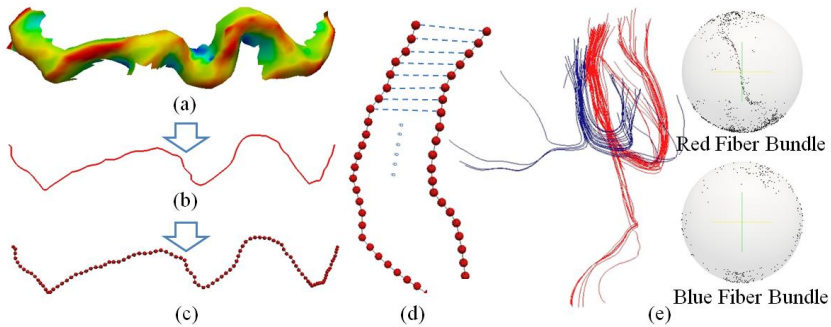

Fig. 2. Illustration of gyral crest line and fiber shape feature. (a) Surface's principal curvature. (b) Gyral crest line. (c) Resampled gyral crest line. (d) Correspondence between gyral nodes on the crest line. (e) Example of fiber bundle shape feature.

2.3   Fiber Shape Feature

We implemented a fiber shape descriptor that is akin to the recently published trace-map model in [7]. Briefly, a fiber tract was divided into overlapping segments.

The principal direction of each segment was calculated using PCA and was represented by a point on a unit-sphere located at origin, as shown in Fig. 2e. The trace maps for each fiber of an ROI are accumulated and a density map was calculated and formulated as a 144-dimensional feature by discretizing the unit sphere and re-arranging the values. Thus, each gyral node is represented by a 144-dimensional trace-map feature, which is quite descriptive in modeling different fiber shape patterns as shown in Fig. 2e. Notably, it has been widely reported that consistent structural connection patterns are predictive of brain function based on the general principle of "connectional fingerprint" introduced in [8], and thus it is expected that the used fiber shape descriptor is predictive of brain function, which will be verified by fMRI data later.

2.4   Adaptive Multi-view Spectral Clustering (AMVSC)

Spectral clustering is a technique that utilizes the properties of the Laplacian of the graph whose edges denote the similarity between the data points. The top c eigenvectors of the symmetric normalized graph Laplacian are the relaxations of the indicator matrix G which assigns each data point in the graph to one of the c clusters. To naturally integrate variable multi-view features of gyral nodes' shape patterns, we propose a unified objective function to simultaneously optimize clustering results of each individual view and their combinations. In other words, we minimize both the summation of the weighted spectral clustering error of each view and the distances between the multi-view clustering indicator matrix and the clustering indicator matrix of each single view. Thus, our multi-view spectral clustering objective function is the following:

$$\min_{G,G^k,\alpha^k} \sum_{k=1}^{v}(\alpha^k)^r Tr(G^{kT}L^kG^k) + \lambda\sum_{k=1}^{v}Tr((G-G^k)^T(G-G^k))$$

$$s.t. \qquad G^TG = I, \qquad \sum_{k=1}^{v}\alpha^k = 1, \qquad \alpha^k \geq 0 \tag{1}$$

where α^k is the normalized weight scalar for k-th view, which can be inferenced from the data directly. $L^k \in R^{n\times n}$ and $G^k \in R^{n\times c}$, are the corresponding Laplacian matrix and clustering indicator matrix of each view respectively $\forall k = 1,2,...,v$. $G \in R^{n\times c}$ is the multi-view clustering indicator matrix that we are interested in. λ is the regularization parameter which controls the tradeoff between the clustering error and consistency and r is the parameter that we use to control the distribution of the weight for each view. Specifically, when $r \to \infty$, we set up equal weight for each view; when $r \to 1$, only one weight is non-zero. Therefore, given the Laplacian matrix of each single view, we utilize Eq. (1) to learn the clustering indicator matrix of each view, clustering indicator matrix for multi-view and weight for each view simultaneously.

2.5   Optimization Algorithms

We will resort to the following approach to solve Eq. (1) alternatively and iteratively. **The first step** is fixing G, G^K, solving α^k. Then we need to solve the next problem,

$$\min_{\alpha^k,\sum_{k=1}^{v}\alpha^k=1,\alpha^k\geq 0} \sum_{k=1}^{v}(\alpha^k)^r p^k \tag{2}$$

Where $p^k = Tr(G^{k^T} L^k G^k)$, and its and its Laplacian function is as follows,

$$\sum_{k}^{v}(\alpha^k)^r p^k - \beta(\sum_{k=1}^{v}\alpha^k - 1) \tag{3}$$

where β is the Laplacian multiplier. Setting the derivative of the Eq. (3) with respective α^k to zero, we have $\alpha^k = (\dfrac{\beta}{rp^k})^{\frac{1}{r-1}}$ Substitute it into the constraint $\sum_{k=1}^{v}\alpha^k = 1$, we obtain $\alpha^k = (rp^k)^{\frac{1}{1-r}} / \sum_{k=1}^{v}(rp^k)^{\frac{1}{1-r}}$.

The second step is fixing α^k, solving G and G^k. Let $\tilde{L} = (\alpha^k)^r L^k$, then the objective function becomes,

$$\min_{G,G^K,G^T G=I} \sum_{k=1}^{v}Tr(G^{k^T}\tilde{L}^k G^k) + \lambda\sum_{k=1}^{v}Tr((G-G^k)^T(G-G^k)) \tag{4}$$

And its Laplacian function is

$$\min_{G,G^k} \sum_{k=1}^{v}Tr(G^{k^T}\tilde{L}^k G^k) + \lambda\sum_{k=1}^{v}Tr((G-G^k)^T(G-G^k)) + Tr(\Lambda^T(G^T G-I)) \tag{5}$$

Where Λ is the Laplacian multiplier. Setting the derivative of Eq. (4) with respective to G^k to zero, then we get: $G^k = \lambda(\tilde{L}^k + \lambda I)^{-1}G$, where $I \in R^{n\times n}$ is the identity matrix. Substitute it into Eq. (4), then obtain the following equivalent problem,

$$\max_{G,G^T G=I} Tr(G^T(\sum_{k=1}^{v}\tilde{L}^k + \lambda I)^{-1})G) \tag{6}$$

which can be tackled by traditional spectral clustering algorithm [9]. Repeating the above two steps alternatively, we iteratively update α^k (thus achieving adaptive MVSC) and G until the objective function converges. We summarize the whole algorithm in **Algorithm 1** [10].

Algorithm 1. The algorithm of AMVSC

Input:

1. The affinity matrix for each view $W^k \in R^{n \times n}, \forall k = 1,2,...,v$.

2. The number of clusters c, the regularization parameter λ and the parameter r.

Output:

1.The cluster indication matrix $G \in R^{n \times c}$.

2.The weight for each view α^k.

Initialization:

1.Set t=0, and initialize the weight for each view, $\alpha_t^1 = \alpha_t^2 = ... = \alpha_t^v = \dfrac{1}{v}$.

Procedure:

1. Calculate the symmetric normalized graph Laplacian for each view $L^k = D^{k-\frac{1}{2}} W^k D^{k-\frac{1}{2}}$, where D^k is the diagonal matrix with $d_{ii}^k = \sum_j w_{ij}^k$.

Repeat

1. Calculate $\widetilde{L}_t^k = (\alpha_t^k)^r L^k$ and calculate $H_t = \sum_{k=1}^{v} (\widetilde{L}_t^k + \lambda I)^{-1}$

2. Update G_{t+1} by solving $\max\limits_{G_t, G_t^T G_t = I} Tr(G_t^T H_t G_t)$ by spectral clustering algorithm [9].

3. Update α_{t+1}^k by $\alpha_{t+1}^k = (rp_{t+1}^k)^{\frac{1}{1-r}} / \sum_{k=1}^{v} (rp_{t+1}^k)^{\frac{1}{1-r}}$, where $p_{t+1}^k = Tr(G_{t+1}^{k}{}^T L^k G_{t+1}^k)$

4. $t = t+1$

Until Converges

3   Experimental Results

3.1   Gyral Parcellation Results

Fig. 4 shows examples of parcellation results for the left pre-central gyrus and right post-central gyrus for 6 subjects, respectively. We set number of segments 10 which gives reasonable parcellation size (neither too small to obtain over segment, nor too large to cover multiple functional regions). By visual examination, the parcellated segments exhibit quite reasonably consistent patterns across different brains. In particular, the results in Figs. 3a and 3b and the results in Figs. 3c and 3d are from two separate datasets, suggesting that the proposed multi-view spectral clustering method really achieved reproducible and consistent parcellation results. To further verify our parcellation result, Figs. 4a and 4b show the fiber shape patterns for all of the gyral segments and Fig. 4c shows the fiber shape patterns of the same segment in 10 brains. It is evident that the fiber shape patterns across different gyral segments are quite different, but the fiber shape patterns of the same gyral segment are quite consistent across individual brains. Quantitatively, the fiber shape similarities within the same corresponding gyral segment are much higher than those between different gyral segments, as shown in Fig. 5. On average, the within-segment similarity is 0.7, while the between-segment similarity is 0.42.
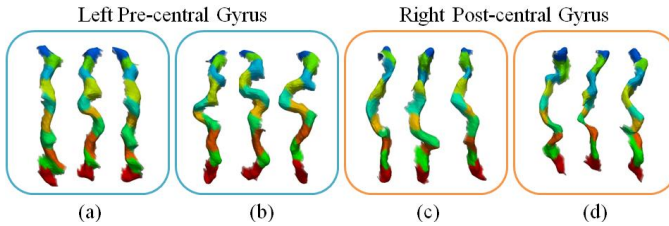
Fig. 3. The parcellation results of the left pre-central gyrus and right post-central gyrus of two datasets. (a) Left pre-central gyrus of subjects in the first dataset. (b) Left pre-central gyrus in the second dataset. (c) Right post-central gyrus in the first dataset. (d) Right post-central gyrus in the second dataset.
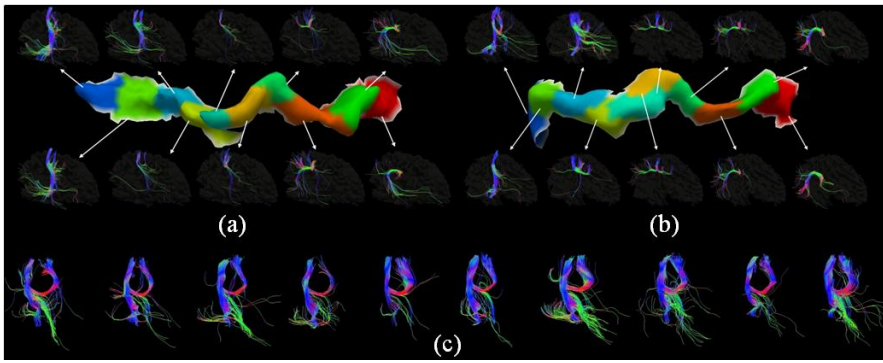

Fig. 4. Examples of fiber shape patterns of each cluster. (a) Fiber shape patterns of the left post-central gyrus of one subject from the first dataset. (b) Fiber shape patterns of the left post-central gyrus of one subject from the second dataset. (c) Examples of fiber shape patterns of the left most segment in (a).
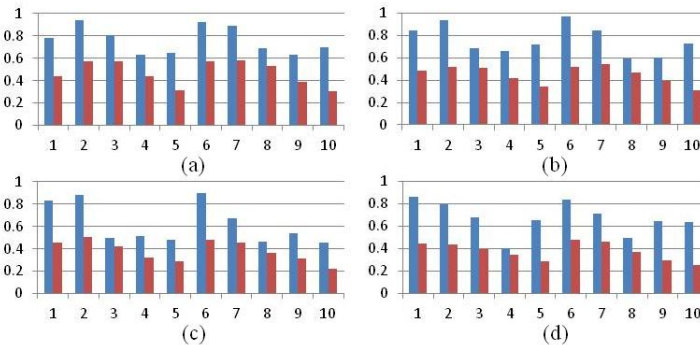

Fig. 5. Average similarity between fiber shape patterns within the same gyral segment in the second dataset. Blue bars: the similarities between subjects of the same segment. Red bars: the similarities between fiber shape patterns of different segments. The horizontal axis is gyral segment ID. (a) Left pre-central gyrus. (b) Right pre-central gyrus. (c) Left post-central gyrus. (d) Right post-central gyrus.
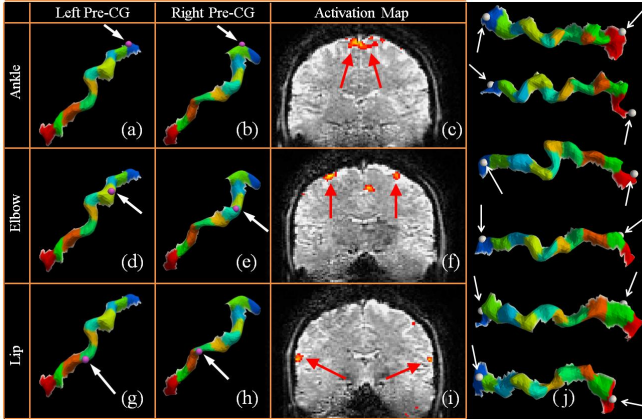
Fig. 6. Evaluation and validation of the pre-central gyrus parcellation by motor task-based fMRI [6] and consistent structural cortical landmarks [11]. (a)-(c) Ankle motor activation. (d)-(f) Elbow motor activation. (g)-(i) Lip motor activation map. (j) Validation by consistent structure cortical landmarks.

3.2   Validation via Motor Task-Based fMRI and Cortical Landmarks

We used benchmark motor task-based fMRI data to validate the functional meanings and correspondences of those structurally parcellated gyral segments. As shown in Figs. 6a-6i, the fMRI-derived activations in responses to ankle, elbow and lip motor tasks are located on the same corresponding segments of the left and right precentral gyri. This result demonstrated that the structurally parcellated gyral segments possess functional correspondences and meanings. In a separate validation study shown in Fig. 6j, two cortical landmarks (highlighted by white arrows) identified in our recent data-drive discovery procedure [11] are located on two corresponding gyral segments across different brains as well, further suggesting that the clustered segments have functional meanings and possess correspondences across individuals.

4   Discussion and Conclusion

This paper presents a novel framework for group-wise consistent parcellation of cortical gyri via multi-view spectral clustering of DTI-derived fiber shape patterns. In this framework, each gyrus's nodes and their associated fiber shape patterns are considered as a separate view in the context of multi-view clustering framework, and gyral parcellation and their group-wise correspondences are achieved simultaneously via the novel multi-view spectral clustering algorithm. Our experimental results have shown that the precentral and postcentral gyri can be clustered into group-wise consistent and meaningful segments. Parts of the clustered results have been validated via benchmark cortical landmarks and task-based fMRI data. In the future, we plan to apply the similar methodology on other cortical gyri and validate the parcellation results via larger scale fMRI studies. The availability of fine-granularity parcellation of the cortex will enable many basic and clinical neuroscience applications [1].

References

1. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis II: inflation, flattening, and a surface-based coordinate system. Neuro Image 9, 195–207 (1999)
2. Shen, D., Davatzikos, C.: HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. IEEE Trans. on Medical Imaging 21(11), 1421–1439 (2002)
3. Li, G., Guo, L., Nie, J., Liu, T.: Automatic Cortical Sulcal Parcellation Based on Surface Principal Direction Flow Field Tracking. Neuroimage 46(4), 923–937 (2009)
4. Meier, J.D., Aflalo, T.N., Kastner, S., Graziano, M.S.: Complex Organization of Human Primary Motor Cortex: A High-Resolution fMRI Study. J. Neurophysiol. 100, 1800–1812 (2008)
5. Li, K., Guo, L., Li, G., Nie, J., Faraco, C., Zhao, Q., Miller, L., Liu, T.: Cortical surface based identification of brain networks using high spatial resolution resting state fMRI data. In: International Symposium of Biomedical Imaging (ISBI), pp. 656–659, 14-17 (2010)
6. Zhu, D., Li, K., Faraco, C., Deng, F., Zhang, D., Jiang, X., Chen, H., Guo, L., Miller, S., Liu, T.: Fine Granularity Parcellation of Gyrus via Fiber Shape and Connectivity Based Features. In: International Symposium of Biomedical Imaging (ISBI), pp. 817–821 (2011)
7. Zhu, D., Li, K., Faraco, C., Deng, F., Zhang, D., Jiang, X., Chen, H., Guo, L., Miller, S., Liu, T.: Optimization of Functional Brain ROIs via Maximization of Consistency of Structural Connectivity Profiles. NeuroImage 59(2), 1382–1393 (2011)
8. Passingham, R.E., Stephan, K.E., Kötter, R.: The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. 3(8), 606-616 (2002)
9. Von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing 17(4), 395–416 (2007)
10. Cai, X., Nie, F., Huang, H., Kamangar, F.: Heterogeneous Image Features Integration via Multi-Modal Spectral Clustering. In: CVPR, pp. 1977–1984 (2011)
11. Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C., Jin, C., Wee, C., Yuan, Y., Lv, P., Yin, Y., Hu, X., Duan, L., Hu, X., Han, J., Wang, L., Shen, D., Miller, L., Li, L., Liu, T.: DICCCOL: Dense Individualized and Common Connectivity-based Cortical Landmarks. Cerebral Cortex (in press, 2012)

Extracting Quantitative Measures from EAP: A Small Clinical Study Using BFOR

A. Pasha Hosseinbor, Moo K. Chung, Yu-Chien Wu, John O. Fleming, Aaron S. Field, and Andrew L. Alexander

University of Wisconsin-Madison, Madison, WI, USA
hosseinbor@wisc.edu

Abstract. The ensemble average propagator (EAP) describes the 3D average diffusion process of water molecules, capturing both its radial and angular contents, and hence providing rich information about complex tissue microstructure properties. Bessel Fourier orientation reconstruction (BFOR) is one of several analytical, non-Cartesian EAP reconstruction schemes employing multiple shell acquisitions that have recently been proposed. Such modeling bases have not yet been fully exploited in the extraction of rotationally invariant q-space indices that describe the degree of diffusion anisotropy/restrictivity. Such quantitative measures include the zero-displacement probability (Po), mean squared displacement (MSD), q-space inverse variance (QIV), and generalized fractional anisotropy (GFA), and all are simply scalar features of the EAP. In this study, a general relationship between MSD and q-space diffusion signal is derived and an EAP-based definition of GFA is introduced. A significant part of the paper is dedicated to utilizing BFOR in a clinical dataset, comprised of 5 multiple sclerosis (MS) patients and 4 healthy controls, to estimate Po, MSD, QIV, and GFA of corpus callosum, and specifically, to see if such indices can detect changes between normal appearing white matter (NAWM) and healthy white matter (WM). Although the sample size is small, this study is a proof of concept that can be extended to larger sample sizes in the future.

1 Introduction

The aim of diffusion-weighted imaging (DWI) is to non-invasively estimate information about the diffusion of water molecules in biological tissues. The most common form of DWI is diffusion tensor imaging (DTI) [4], which is a good model of diffusion-weighted signal behavior at low levels of diffusion weighting. Rotationally invariant measures can be derived from the eigenvalues of the diffusion tensor, including fractional anisotropy (FA) and mean diffusivity (MD) [5], that have proven clinical value. However, DTI is limited by the Gaussian assumption, which is invalid at higher levels of diffusion weighting ($b > 2000$ s/mm^2) and its inability to resolve multiple fiber orientations within a voxel [12].

In order to recover complex white matter (WM) geometry, high angular resolution diffusion imaging (HARDI) [12], which reduces the diffusion signal sampling to a single sphere (i.e. single level of diffusion weighting) within q-space,

was proposed. Many HARDI techniques [7,11] seek to extract the orientation distribution function (ODF), a probability density function describing the angular distribution of water molecules during diffusion. However, the ODF only retrieves the angular content of the diffusion process.

The ensemble average propagator (EAP) provides the full information about the diffusion process in the tissue because it captures both the radial and angular information contained in the diffusion signal. The ODF is simply an angular feature of the EAP. Unlike the diffusion tensor, the EAP profiles illustrate and recover crossing fibers. The significance of the EAP in diffusion MRI has led to many reconstruction algorithms being proposed, some numerically based such as diffusion specturm imaging (DSI) [13] and hybrid diffusion imaging (HYDI) [16], and some analytically based such as diffusion propagator imaging (DPI) [9], spherical polar Fourier imaging (SPFI) [3,8], and Bessel Fourier orientation reconstruction (BFOR) [10].

With respect to analytical EAP reconstruction methods, one valuable though overlooked use is in extracting rotationally invariant quantitative measures from them. High angular resolution analogues of quantitative DTI indices such as generalized fractional anisotropy (GFA) [11] & mean squared displacement (MSD) [2,16] and other q-space metrics like zero-displacement probablity (Po) [2] & q-space inverse variance (QIV) [17] are all scalar features of the EAP. Analytical representations of the EAP (and hence diffusion signal) facillitate either analytic computation of such features or numerical efficiency in estimating them. HYDI has already been used to numerically estimate Po, MSD, and QIV [17].

In this paper, we derive analytical expressions for Po, MSD, & QIV using BFOR, and introduce an EAP-based definition of GFA. These quantitative measures are then utilized in a HYDI-acquired clinical dataset, comprising a healthy control group and multiple sclerosis (MS) patients, to see if they detect any differences in the corpus callosum between the normal appearing white watter (NAWM) of MS patients and healthy WM.

2   Theory

Let $P(\mathbf{p})$ and $E(\mathbf{q})$ be the EAP and normalized q-space diffusion signal, respectively. We denote $\mathbf{q} = q\ \mathbf{u}(\theta, \phi)$ and $\mathbf{p} = p\ \mathbf{r}(\theta', \phi')$, where $\mathbf{u}$ and $\mathbf{r}$ are 3D unit vectors. Under the narrow pulse assumption, $E(\mathbf{q})$ and $P(\mathbf{p})$ are Fourier Transform (FT) pairs [6]:

$$P(\mathbf{p}) = \int E(\mathbf{q}) e^{-2\pi i \mathbf{q} \cdot \mathbf{p}} d^3 \mathbf{q} \tag{1}$$

The BFOR signal basis and EAP are, respectively,

$$E(\mathbf{q}, t) = \sum_{n=1}^{N} \sum_{j=1}^{R} C_{nj} e^{\frac{-\alpha_{nl(j)}^2 t}{\tau^2}} j_{l(j)}\left(\frac{\alpha_{nl(j)} q}{\tau}\right) Y_j(\mathbf{u}) \tag{2}$$

and

$$P(\mathbf{p}, t) = 2\tau\sqrt{2\pi^3} \sum_{n=1}^{N} \sum_{j=1}^{R} (-1)^{\frac{l(j)}{2}} C_{nj} e^{\frac{-\alpha_{nl(j)}^2 t}{\tau^2}} Y_j(\mathbf{r}) \frac{\sqrt{\alpha_{nl(j)}} J_{l(j)-1/2}(\alpha_{nl(j)}) j_{l(j)}(2\pi\tau p)}{\left(4\pi^2 p^2 - \frac{\alpha_{nl(j)}^2}{\tau^2}\right)},$$
(3)

where $e^{\frac{-\alpha_{nl(j)}^2 t}{\tau^2}}$ is the smoothening term, C_{nj} are the expansion coefficients, and $\alpha_{nl(j)}$ is n^{th} root of lth order spherical Bessel function of first kind j_l [10].

2.1   Rotationally Invariant Quantitative q-Space Indices

$Po = P(\mathbf{p} = 0)$ is the probability density of water molecules that minimally diffuse within the diffusion time [2,16], and hence a measure of restricted diffusion. In a healthy adult brain, Po is greater in white matter (WM) than gray matter (GM) because WM has more restricting barriers including multi-layer myelin sheaths, axonal membranes, and microtubules. Several studies have shown Po to be sensitive to brain pathology, and suggesting that changes in myelin are the primary mechanism for differences in Po [1,18].

Po can be evaluated either numerically or analytically. The authors in [17] computed Po by numerically summing the normalized diffusion signal $E(\mathbf{q})$ over all diffusion measurements in q-space, and then correcting the sum by the sampling density. Analytical formulations of Po were derived for the SPFI and DPI signal bases [8,9]. The BFOR Po can be computed analytically by evaluating Eq. (3) at $p = 0$[1]:

$$Po = 2\sqrt{\pi}\tau^3 \sum_{n=1}^{N} C_{n1} \frac{(-1)^{n+1}}{\alpha_{n0}^2}$$
(4)

The MSD, which we will denote as $\langle p^2 \rangle$, is simply the second moment of the EAP [16]: $\langle p^2 \rangle = \int p^2 P(\mathbf{p}) d^3 \mathbf{p}$. It is related to the MD, which in the case of Gaussian diffusion is given by the well-known Einstein relation $\langle p^2 \rangle = 6(\Delta - \delta/3)\text{MD}$. Thus far, an analytical formulation of MSD exists only within the DTI framework. It is calculated numerically in q-space imaging, either by extracting the full width at half maximum of the EAP [2] or taking the geometrical mean of the diffusion signal over all directions on a HYDI shell [17]

A general relationship between the MSD and q-space diffusion signal has not yet been formulated to the authors' knowledge. Such a relationship is derived in the **Supplementary Section**:

$$\langle p^2 \rangle = \frac{-1}{4\pi^2} \bigtriangledown^2 E(\mathbf{q})|_{\mathbf{q}=0}$$
(5)

According to Eq. (5), DPI, which models the diffusion signal as $\bigtriangledown^2 E(\mathbf{q}) = 0$, predicts the MSD to be zero which is unrealistic. Using Eq. (5), an analytic MSD

---

[1] For all derivations, see http://brainimaging.waisman.wisc.edu/~ameer/Suppl.pdf

expression can be computed for BFOR because the BFOR signal basis, Eq. (2), is an eigenfunction of the Laplacian operator. Hence, it can be shown that

$$\langle p^2 \rangle_{BFOR} = \frac{1}{8\pi^{\frac{5}{2}}\tau^2} \sum_{n=1}^{N} C_{n1} \alpha_{n0}^2 \tag{6}$$

The MSD measure is quite sensitive to noise [16]. The authors in [17] proposed an alternative measure to MSD called the QIV, which is a pseudo-diffusivity measure. Mathematically, the QIV is defined as $QIV = \left[\int q^2 E(\mathbf{q}) d^3\mathbf{q} \right]^{-1}$. The QIV is not an arbitrary meausre, but related to the EAP in a manner analogous to which the MSD is related to the diffusion signal-in **Suppl. Section**, we will show that $QIV^{-1} = \frac{-1}{4\pi^2} \triangledown^2 P(\mathbf{p})|_{\mathbf{p}=0}$. The QIV within the BFOR framework is

$$QIV_{BFOR} = \frac{1}{2\sqrt{\pi}\tau^5 \sum_{n=1}^{N}(-1)^n C_{n1} \frac{(6-\alpha_{n0}^2)}{\alpha_{n0}^4}} \tag{7}$$

Tuch in [11] introduced the concept of GFA and defined it as std(ODF)/ rms(ODF). Since ODF is only a feature of the EAP, the subsequent GFA map is derived soley from the angular content of the diffusion profile. Incorporating both the angular and radial contents of the diffusion profile into the definition of GFA will result in a radial dial of GFA maps, illustrating how anisotropy varies with diffusion displacement p. Therefore, we define a new GFA:

$$GFA(p = p_o) = \frac{std\left[P(p = p_o, \mathbf{r})\right]}{rms\left[P(p = p_o, \mathbf{r})\right]} \tag{8}$$

Another advantage of Eq. (8) is that it is better suited for multiple diffusion weighted MR experiments, unlike Tuch's definition, which is HARDI-based.

3   Materials and Methods

The *in vivo* dataset uses a hybrid, non-Cartesian sampling scheme [16], shown in Table 1. Since EAP reconstruction is sensitive to angular resolution, the number of encoding directions is increased with each shell to increase the angular resolution with the level of diffusion weighting. The number of directions in the outer shells were increased to better characterize complex tissue organization.

HYDI was performed on five MS patients and four healthy volunteers using a 3.0 T GE-SIGNA whole body scanner. MR parameters were TE = 99 ms, TR 2300 ms, FOV = 24 cm, matrix = 96 x 96, voxel size = 2.5 x 2.5 mm^2, 15 slices with slice thickness = 5 mm, and scan time = 10 min. Diffusion parameters were $\Delta = 45$ ms, $\delta = 34$ ms, field of view of the diffusion displacement space FOV_p = $(1/\Delta q)$ = 71.4 μm, and resolution of the diffusion displacement space Δp = $(1/2q_{max})$ = 7.1 μm [6].

DTI analysis was performed using the data in the second HYDI shell, in order to obtain the FA and MD, with the FSL software package [15]. BFOR was then used to compute Po, MSD, QIV, and GFA, with model parameters
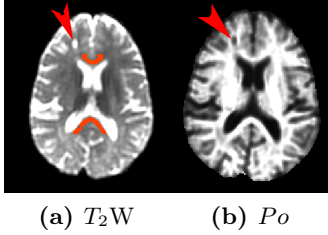
(a) T_2W   **(b)** Po

Fig. 1. Axial slice of T_2W and corresponding BFOR Po map illustrating corpus callosum ROI (red) and MS lesion (red arrow).

**Table 1.** HYDI Encoding Scheme for Human Dataset

Shell	Ne	q (mm^{-1})	Δq (mm^{-1})	b (s/mm^2)
	2	0		0
1st	6	14	14	260
2nd	21	28	14	1040
3rd	24	42	14	2340
4th	24	56	14	4160
5th	50	70	14	6500

set to $\{L = 4, N = 6, \tau = 84 \text{ mm}^{-1}, \lambda_l = 10^{-6}, \lambda_n = 10^{-6}, t = 0\}$. Using the T_2W ($b = 0$ volume) and BFOR Po maps as references, ROIs of the genu and splenium of corpus callosum were then manually drawn for each subject, as shown in Fig. 1, which were also applied to the other quantitative maps. An unpaired two-sample t-test (one-tailed), assuming unequal variances, was then used to test whether the mean value of each index in the corpus callosum for the NAWM group was lower (FA, GFA & Po) or higher (DTI/BFOR MSD & QIV) than those from control group at 0.05 level.

4   Results

Fig. 4 displays axial slices of the BFOR computed Po, QIV, & MSD indices. Note that the QIV exhibits GM/WM contrast, unlike MSD. Within the CSF regions in QIV map, some voxels were zeroed out because they blew up upon the division operation in computing QIV. Fig. 3 shows axial slices of the GFA estimated at $p = 5$, 10, and 15 μm, illustrating how the anisotropy of different WM regions, such as the corpus callosum and capsules, varies with diffusion displacement p. CSF regions in the GFA map at $p = 15$ μm are more noisy than at 5 & 10 μm.

Fig. 2 displays the mean and standard deviation of GFA, MSD, QIV, Po, FA, and DTI MSD for each subject. The t-test yielded a statistically significant p-value between the mean value of each index in NAWM and healthy WM at 0.05 level for GFA(5) & GFA(10) and Po, implying a reduction in GFA and Po of NAWM in corpus callosum with respect to controls. Such findings are consistent with previous DTI [14] and q-space [1] MS studies that showed significant reductions in FA and Po of NAWM with respect to controls, respectively. Although the DTI FA was also statistically significant, the p-values for GFA(5) & GFA(10) are much smaller than for FA, suggesting that GFA may be more sensitive to pathologically induced changes in WM than normal FA. The BFOR MSD was not a statistically significant indicator of pathological changes in WM, which goes against the results of [1,14] that showed MD/MSD to be significantly
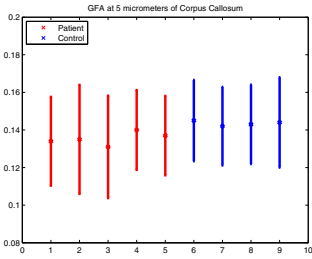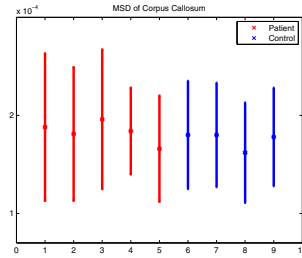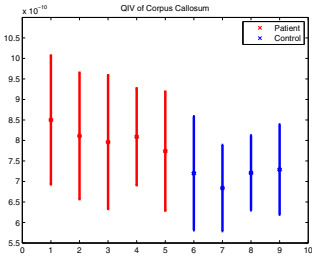
(a) GFA(5); p-value=0.0018
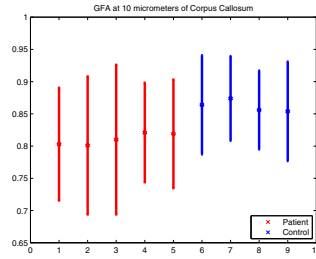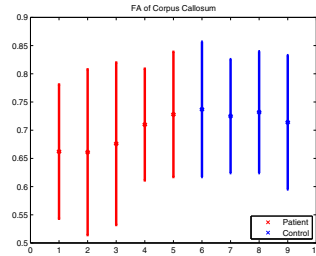
(b) GFA(10); p-value=4.8e-5
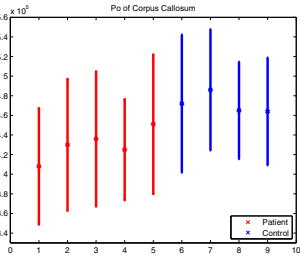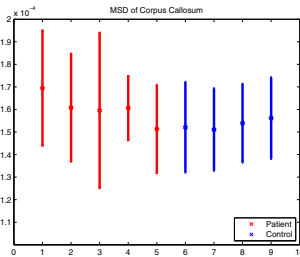
(c) *Po*; p-value=0.0011

(d) BFOR MSD; p-value=0.13

(e) QIV; p-value=3.0e-4

(f) DTI FA; p-value=0.020

(g) DTI MSD; p-value=0.035

Fig. 2. Mean and stdv. of measures in corpus callosum for each subject, with the p-value of the unpaired two-sample t-test (one-tailed)
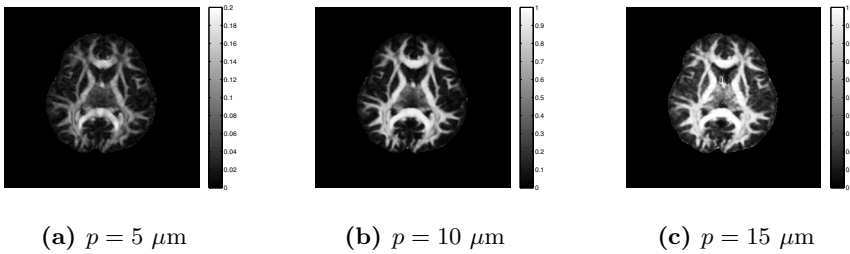
(a) $p = 5 \ \mu$m   **(b)** $p = 10 \ \mu$m   **(c)** $p = 15 \ \mu$m

Fig. 3. Axial slices of BFOR GFA maps at $p = 5$, 10, and 15 μm for a control

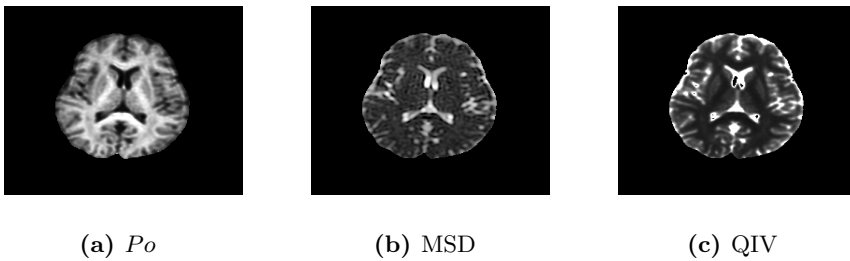**(a)** Po **(b)** MSD   **(c)** QIV

Fig. 4. Axial slices of BFOR estimated Po, MSD, and QIV maps for a control

higher in NAWM with respect to controls. The DTI MSD, however, was found to be statistically significant, and the disparity in results between the DTI and BFOR MSD may be due to the high b-value (BFOR) MSD being very sensitive to noise [16]. The QIV, however, was found to be significantly higher in NAWM with respect to controls, and both it and GFA(10) had the highest statistical significances among all metrics. In general, the BFOR computed measures validate the main finding of [1], being that q-space indices suggest abnormalities in the MS brain are not only confined to hyerintense lesions visible in T_2 images, but may also affect the surrounding NAWM.

5 Conclusion

This is the first study to date to utilize an analytical, hybrid, and non-Cartesian EAP framework for the computation of rotationally invariant quantitative measures in a clinical dataset. Although the study was limited by the small sample size, it demonstrates the potential that EAP-derived q-space indices have in assessing brain pathology. In the future, the same study should be repeated using a larger sample size, with measurements being made in other WM regions in addition to the corpus callosum. Future work also includes estimating the axial and radial diffusivities using an analytical EAP framework.

References

1. Assaf, Y., Ben Bashat, D., Chapman, J., Peled, S., Biton, I.E., Kafri, M., Segev, Y., Hendler, T., Korczyn, A.D., Graif, M., Cohen, Y.: High b-value q-space analyzed diffusion-weighted MRI: application to multiple sclerosis. Magn. Reson. Med. 47, 115–126 (2002)
2. Assaf, Y., Mayk, A., Cohen, Y.: Displacement imaging of spinal cord using q-space diffusion-weighted MRI. Magn. Reson. Med. 44, 713–722 (2000)
3. Assemlal, H.E., Tschumperlé, D., Brun, L.: Efficient and robust computation of PDF features from diffusion MR signal. Med. Image Anal. 13, 715–729 (2009)
4. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophysical Journal 66, 259–267 (1994)
5. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J. Magn. Reson. 111, 209–219 (1996)
6. Callaghan, P.T.: Principles of Nuclear Magnetic Resonance Microscopy. Oxford University Press, Oxford (1991)
7. Canales-Rodriguez, E.J., Melie-Garcia, L., Iturria-Medina, Y.: Mathematical description of q-space in spherical coordinates: exact q-ball imaging. Magn. Reson. Med. 61, 1350–1367 (2009)
8. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: Model-Free and Analytical EAP Reconstruction via Spherical Polar Fourier Diffusion MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 590–597. Springer, Heidelberg (2010)
9. Descoteaux, M., Deriche, R., LeBihan, D., Mangin, J.F., Poupon, C.: Multiple q-shell diffusion propagator imaging. Med. Image Anal. 15, 603–621 (2011)
10. Hosseinbor, A.P., Chung, M.K., Wu, Y.-C., Alexander, A.L.: Bessel Fourier Orientation Reconstruction: An Analytical EAP Reconstruction Using Multiple Shell Acquisitions in Diffusion MRI. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 217–225. Springer, Heidelberg (2011)
11. Tuch, D.S.: Q-ball imaging. Magn. Reson. Med. 52, 1358–1372 (2004)
12. Tuch, D.S., Reese, T.G., Wiegell, M.R., Makris, N., Belliveau, J.W., Weeden, V.J.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magn. Reson. Med. 48, 577–582 (2002)
13. Weeden, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magn. Reson. Med. 54, 1377–1386 (2005)
14. Werring, D.J., Clark, C.A., Barker, G.J., Thompson, A.J., Miller, D.H.: Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis. Neurology 52, 1626–1632 (1999)
15. Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M.: Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, 173–186 (2009)
16. Wu, Y.C., Alexander, A.L.: Hybrid diffusion imaging. NeuroImage 36, 617–629 (2007)
17. Wu, Y.C., Field, A.S., Alexander, A.L.: Computation of diffusion function measures in q-space using magnetic resonance hybrid diffusion imaging. IEEE Transac. Med. Imaging 27, 858–865 (2008)
18. Wu, Y.C., Field, A.S., Duncan, I.D., Samsonov, A.A., Kondo, Y., Tudorascu, D., Alexander, A.L.: High b-value and diffusion tensor imaging in a canine model of dysmyelination and brain maturation. NeuroImage 58, 829–837 (2011)

Sparse DSI: Learning DSI Structure
for Denoising and Fast Imaging

Alexandre Gramfort[1,2], Cyril Poupon[2], and Maxime Descoteaux[3]

[1] Parietal Team, INRIA Saclay-Île-de-France, Saclay, France
alexandre.gramfort@inria.fr
[2] CEA, DSV, I²BM, Neurospin bât 145, 91191 Gif-Sur-Yvette, France
[3] SCIL, Computer Science department, Université de Sherbrooke

Abstract. Diffusion spectrum imaging (DSI) from multiple diffusion-weighted images (DWI) allows to image the complex geometry of water diffusion in biological tissue. To capture the structure of DSI data, we propose to use sparse coding constrained by physical properties of the signal, namely symmetry and positivity, to learn a dictionary of diffusion profiles. Given this estimated model of the signal, we can extract better estimates of the signal from noisy measurements and also speed up acquisition by reducing the number of acquired DWI while giving access to high resolution DSI data. The method learns jointly for all the acquired DWI and scales to full brain data. Working with two sets of 515 DWI images acquired on two different subjects we show that using just half of the data (258 DWI) we can better predict the other 257 DWI than the classic symmetry procedure. The observation holds even if the diffusion profiles are estimated on a different subject dataset from an undersampled q-space of 40 measurements.

1 Introduction

Diffusion-weighted imaging offers a way to non-invasively image the diffusion of water molecules in biological tissue. The ensemble average propagator (EAP) formalism provides a powerful framework to describe and predict the diffusion behavior of water molecules in complex materials. Under the narrow pulse assumption, there is a Fourier relationship between the measured DWI signal and diffusion propagator, $P(\mathbf{R})$,

$$P(\mathbf{R}) = \int_{\mathbf{q} \in \mathbb{R}^3} E(\mathbf{q}) e^{-2\pi i \mathbf{q} \cdot \mathbf{R}} d\mathbf{q}, \tag{1}$$

with $E(\mathbf{q}) = S(\mathbf{q})/S_0$, where $S(\mathbf{q})$ is the diffusion signal measured at position $\mathbf{q}$ in q-space, and S_0 is the baseline image acquired without any diffusion sensitization ($q = 0$). We denote $q = |\mathbf{q}|$ and $\mathbf{q} = q\,\mathbf{u}$, $\mathbf{R} = r\,\mathbf{r}$, where $\mathbf{u}$ and $\mathbf{r}$ are 3D unit vectors. The norm of the wave vector, $\mathbf{q}$, is related to the diffusion weighting factor (the b-value), $b = 4\pi^2 q^2 \tau$, where τ is the effective diffusion time.

Diffusion Spectrum Imaging (DSI) reconstruct the EAP. The original DSI protocol [1] measures $S(\mathbf{q})$ on a Cartesian grid inscribed in a sphere of radius

5, resulting in 515 q-space discrete measurements $S(\mathbf{q})$. Then, a Hanning filter is applied on the raw $S(\mathbf{q})$'s to reduce truncation errors and boundary artifacts before a simple 3D Fast Fourier Transform (FFT) is applied to recover the EAP at every imaging voxel. Finally, the diffusion orientation distribution function (ODF), Ψ, can be extracted by numerically computing the radial integral over $r \in [0, 5]$, as $\Psi(\mathbf{u}) = \int_0^5 P(r\mathbf{u})r^2 dr$. DSI is a long acquisition (approximately 1h45 min for a full brain with 2mm isotropic voxels). Because diffusion is symmetric, one can cut the acquisition time in half if only one hemi-plane is acquired, resulting in 258 directions [2]. DSI has regained popularity in the last years because successful connectomics studies [2] and the human brain connectome project has brought back DSI as the central DWI protocol before fiber tractography.

As any experimental data, DSI data is corrupted by noise, especially because of the large b-values used in the protocol. At these larger b-values, most of the signal is near the noise floor (see Fig. 3). A first challenge is therefore to be able to improve the quality of DSI with denoising algorithms. Another key challenge is the ability to reduce the acquisition time while offering high resolution data required to estimate complex fiber crossings. Hence the goals of this paper are twofold: 1) Denoising of DSI data. 2) Fast DSI acquisition for clinical applications. In particular, can we subsample q-space while keeping good high spectral resolution, i.e. perform DSI at the price of HARDI?

The intuition behind this paper is that the DSI acquisition on 258 points hemi-plane or the full 515 sampling contains redundant information that one can learn to denoise or accelerate DSI acquisition. We propose to use sparse coding to estimate a dictionary of prototypical DSI profiles that well model the structure of the DSI signal. We show that the estimated dictionary of DSI profiles captures the geometry of white matter brain structures and can thus be used to improve the data quality while keeping a small sampling of the q-space. It can be used for 2 things: i) intra-subject studies for denoising purposes and ii) inter-subject studies by using a lower DSI sampling acquisitions to recover the full DSI using a learned dictionary of DSI profiles.

Notations We write vectors in bold, $\mathbf{a} \in \mathbb{R}^n$, matrices with capital bold letters, $\mathbf{A} \in \mathbb{R}^{n \times n}$. A scalar a is positive if $a \in \mathbb{R}_+$. We denote $\|\mathbf{A}\|_{\mathrm{Fro}}$ the Frobenius norm, $\|\mathbf{A}\|_{\mathrm{Fro}}^2 = \sum_{i,j=1}^n \mathbf{A}_{ij}^2$, and $\|\mathbf{A}\|_1 = \sum_{i,j=1}^n |\mathbf{A}_{ij}|$ the ℓ_1 norm. Column i of a matrix is written $\mathbf{A}^i$. If $\mathcal{I}$ is a list of $|\mathcal{I}|$ indices, $\mathbf{A}_{\mathcal{I}}$ is the matrix $\mathbf{A}$ restricted to the rows in $\mathcal{I}$. $\mathbf{I}$ stands for the identity matrix. Quantities estimated from the data are written $\widehat{\mathbf{A}}$. A matrix with non-negative elements is denoted $\mathbf{A} \geqslant 0$.

2 Learning a Dictionary of DSI Profiles with Sparse Coding

Let $\mathbf{S} \in \mathbb{R}_+^{d \times p}$ denote acquired DWI at p voxels over a set of d directions ($d = 258$ in the results section below). We consider the following generative model for the DWI data at voxel i: $\mathbf{S}^i = \mathbf{D}\mathbf{W}^i + \mathbf{e}^i$, where $\mathbf{D} \in \mathbb{R}_+^{d \times k}$ is a dictionary of DSI profiles and $\mathbf{W} \in \mathbb{R}_+^{k \times p}$ are the coefficients of the data decomposition over the

dictionary. The integer k is the size of the dictionary. The noise $\mathbf{e}^i \in \mathbb{R}^d_+$ is known to have a Rician distribution and non-central χ-distribution when using parallel imaging [3,4]. However, for the current contribution, it will be assumed Gaussian with mean $\mu \in \mathbb{R}^d$ and diagonal covariance $\mathbf{\Sigma} = \mathrm{diag}((\sigma^2_j)_{j=1,...,d}) \in \mathbb{R}^{d \times d}$. We denote it $\mathbf{e}^i \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. This modeling assumption is valid for images with SNR above 4 [3], which is the case for the DSI data used in this paper.

The estimation procedure detailed below requires that the additive noise to be Gaussian with unit variance. To meet this constraint, we define the whitened data $\mathbf{S}^i_w = \mathbf{\Sigma}^{-1/2}(\mathbf{S}^i - \mu)$ so that $\mathbf{S}^i_w \sim \mathcal{N}(\mathbf{D}_w\mathbf{W}^i, \mathbf{I})$ where $\mathbf{D}_w = \mathbf{\Sigma}^{-1/2}\mathbf{D}$. In practice, μ and $\mathbf{\Sigma}$ are estimated from voxels with weak diffusion (e.g. the skull).

In order to learn the dictionary $\mathbf{D}$ one needs to set priors on both $\mathbf{D}$ and $\mathbf{W}$. The positivity of DWI data imposes that $\mathbf{D} \geqslant 0$ and $\mathbf{W} \geqslant 0$. The dictionary is a good model for the entire white matter but the data at a given voxel should only be formed by a linear combination of a few DSI profiles: $\mathbf{W}$ should be sparse, i.e. contain many zeros. This leads to the following minimization:

$$(\widehat{\mathbf{D}}_w, \widehat{\mathbf{W}}) = \underset{\mathbf{D},\mathbf{W}}{\mathrm{argmin}} \frac{1}{2d}\|\mathbf{S}_w - \mathbf{D}_w\mathbf{W}\|^2_{\mathrm{Fro}} + \lambda\|\mathbf{W}\|_1 \tag{2}$$

$$\text{s.t. } \|\mathbf{D}^k\|^2_2 \leqslant 1, \ \mathbf{D} \geqslant 0, \ \mathbf{W} \geqslant 0$$

The parameter λ balances the reconstruction error and the ℓ_1 regularization term which promotes sparse coefficients $\mathbf{W}$. Columns of $\mathbf{D}$ have unit norm to avoid scaling ambiguity. Following [5], we use an online cyclic descent to minimize (2). The estimated dictionary $\widehat{\mathbf{D}}$ is then given by $\widehat{\mathbf{D}} = \mathbf{\Sigma}^{1/2}\widehat{\mathbf{D}}_w$.

Denoising and subsampling. Given a learned dictionary $\mathbf{D}$, one can estimate the coefficients of a decomposition for a new dataset, with possibly less directions than the original dataset used to learn $\mathbf{D}$, i.e. using only a limited set of rows of $\mathbf{D}$. Let us denote $\mathcal{I}$ a list of sampling directions ($|\mathcal{I}| \leqslant d$). Given a set of whitened subsampled data $\mathbf{S}_{\mathcal{I}w}$, the coefficients $\mathbf{W}$ can be obtained by solving:

$$\widehat{\mathbf{W}} = \underset{\mathbf{W}}{\mathrm{argmin}} \frac{1}{2|\mathcal{I}|}\|\mathbf{S}_{\mathcal{I}w} - \mathbf{D}_{\mathcal{I}w}\mathbf{W}\|^2_{\mathrm{Fro}} + \delta\|\mathbf{W}\|_1 \qquad \text{s.t. } \mathbf{W} \geqslant 0$$

where $\delta > 0$. The full signal can then be obtained as: $\widehat{\mathbf{S}} = \mathbf{D}\widehat{\mathbf{W}} \in \mathbb{R}^{d \times p}$.

Model selection and choice of parameters. The estimation procedure involves a few parameters, λ to learn a dictionary and δ to estimate the coefficients given the dictionary. In order to tune these parameters we use cross-validation exploiting the symmetry of the signal.

Given half of the directions $\mathcal{H}$ (d=258 directions), a good model should be able to predict the other half by applying a simple symmetry to the data. This observation also allows us to estimate a full dictionary of $2d-1 = 515$ directions using only half of the data. The minimization of (2) restricted to the data $\mathbf{S}_{\mathcal{H}}$ gives $\widehat{\mathbf{D}}_{\mathcal{H}} \in \mathbb{R}^{d \times k}$ which leads to $\widehat{\mathbf{D}} \in \mathbb{R}^{2d-1 \times k}$ by applying a symmetry.
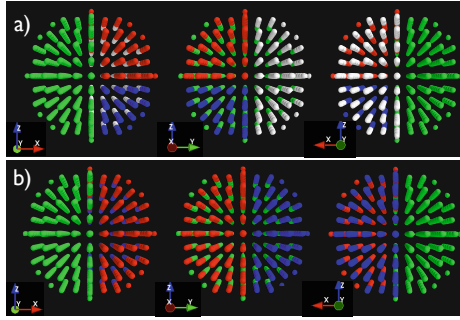
Fig. 1. In green are the 258 directions $\mathcal{H}$ used to estimate a dictionary. a) For intra-subject cross-validation, in red are the directions $\mathcal{L}$ used to estimate the coefficients and in blue the left out directions $\mathcal{T}$ used to evaluate the model parameters. The white directions $\mathcal{V}$ are only used for validation. b) For inter-subject cross-validation the same color code applies. There is no validation set in this case, as the validation data is obtained from the other subject.

In order to assess the quality of the model without overfitting, the model selection involves two other sets of directions: $\mathcal{L}$ to learn $\widehat{\mathbf{W}}$ and $\mathcal{T}$ to test the reconstruction error. The best parameters minimize this error. It is quantified with the average root mean square error (RMSE): RMSE $= \frac{1}{p|\mathcal{T}|}\|\mathbf{S}_{\mathcal{T}} - \widehat{\mathbf{D}}_{\mathcal{T}}\widehat{\mathbf{W}}\|_{\text{Fro}}^2$. This procedure is a principled way of choosing the parameters. Once they are set, the model estimated can be tested on new DWI $\mathcal{V}$ for validation. As $\mathcal{V}$ has not been used for training, RMSE obtained by different algorithms can be compared.

In the following experiments, the parameter λ was set in a range of 5 values $(1, 0.1, 0.01, 0.001)$ and δ in a logarithmic grid of 15 values between 0 and $1e^{-6}$. In order to quantify the performance of our method, we use as baseline the solution that consists in applying a simple symmetry to the data, as done classically [2]. We denote RMSE$_{sym}$ the error obtained. We report the quality of our solution as a ratio between the two quantities: $\rho_{RMSE} = \text{RMSE}_{sym}/\text{RMSE}$. A ratio above 1 indicates an improvement with respect to a symmetrization.

The following results involve two setups. An intra-subject denoising procedure and an inter-subject procedure where a dictionary learned on a subject is used to estimate a full resolution DSI dataset for a new subject from only a few DWI. See Fig. 1 for details on $\mathcal{H}, \mathcal{L}, \mathcal{T}$ and finally $\mathcal{V}$ used to estimate ρ_{RMSE}.

The data consists of 2 subjects. A standard DSI acquisition mimicking [1] was done with isotropic 2 mm spatial resolution and $d = 515$ DWI were acquired sampling the q-space on a cubic lattice within the sphere of radius 5, TE/TR=147ms/11.5s, BW=1680Hz/pixel, 96x96 matrix, and 60 axial slices with a parallel reduction factor of 2, δ and Δ were 41 and 45 ms, resulting in a maximum q-value of $q_{\max} = 70.4$ mm^{-1}, $b_{\max} = 6000$ s/mm^2. The SNR of the $b = 0$ image was 36 and the SNR of the DWI for the $b = 960, 3360$, and 6000 s/mm^2 datasets were estimated to 12, 7.5, and 6.5 respectively. For the rest of this paper, acquisition of 258 directions with simple symmetry will

be abbreviated HALF, as opposed to the FULL acquisition. Different symmetry completion procedure are also included for comparison, either using Gaussian smoothing with $\sigma = 0.35$ (optimal σ in our experiments) and state-of-the-art non-Local (NL) means denoising [6]. Finally, ODFs are computed as described above and visualized as deformed spheres with the radius proportional to $\Psi(\mathbf{u})$.

3   Results

Figure 2 shows the ODFs of the learned dictionaries with 100 atoms on both subjects. Atoms are ordered from left to right, starting at the bottom left corner based on the variance they explain on the data. We see that most important atoms are isotropic profiles and several single fiber structures. After approximately 30 atoms, crossing profiles appear. At the end of the dictionary, more complex ODF profiles are also present. This behavior of the learned dictionary is similar if we increase its size k.

Intra-subject denoising Table 1 shows how sparse DSI reconstruction is able to accurately reconstruct the un-measured 257 directions. It has lower RMSE than the usual symmetry and, Gaussian and NL means denoising. Increasing the number of atoms in the dictionary only slightly improves the accuracy on subject 2. Moreover, Fig. 3 confirms that denoising, in general, improves the raw DSI data. However, it can be seen that NL means and Gaussian denoising seem to over-smooth and blur the structure of the raw data, as opposed to sparse DSI that appear to denoise but also enhance structure. Finally, Fig. 4 overlays ODFs in a zoom region of this slice, corresponding to the centrum semiovale where corpus callosum (CC) crosses with the corticospinal tract (CT) and superior longitudinal fasciculus (SLF). Single, two and three fiber crossings can be seen. One can appreciate how sparse DSI is able to recover ODF profiles as sharp as the FULL raw and NL means/Gaussian denoised DSI.

Inter-subject undersampling One can push sparse DSI and attempt to perform DSI estimation and ODF reconstruction from undersampled q-space data. The compressed sensing literature teaches us that the "sensing" strategy is crucial for optimal reconstructions. It is beyond the scope of this paper to explore optimal undersampling strategies. Here, we undersampled 1 measurement out of N from the Cartesian direction indices, which preserves a uniform Cartesian sampling.
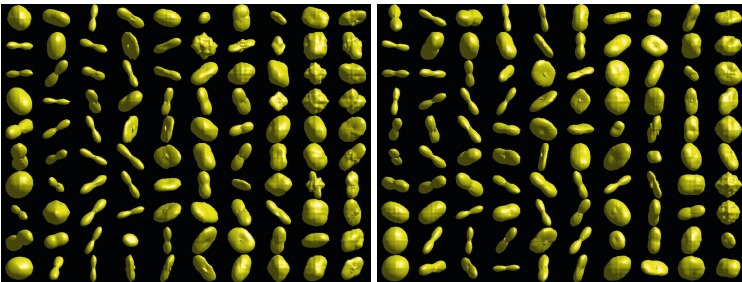

Fig. 2. ODFs computed from the learned dictionaries on the 2 subjects (100 atoms and inter-subject cross-validation). Left (resp. right) is for subject 1 (resp. 2).

Table 1. Intra-subject denoising. ρ_{RMSE} between simple DSI symmetry, Gaussian smoothing, NL means and sparse DSI denoising. Sparse DSI reconstruction gives the best performance on the validation data.

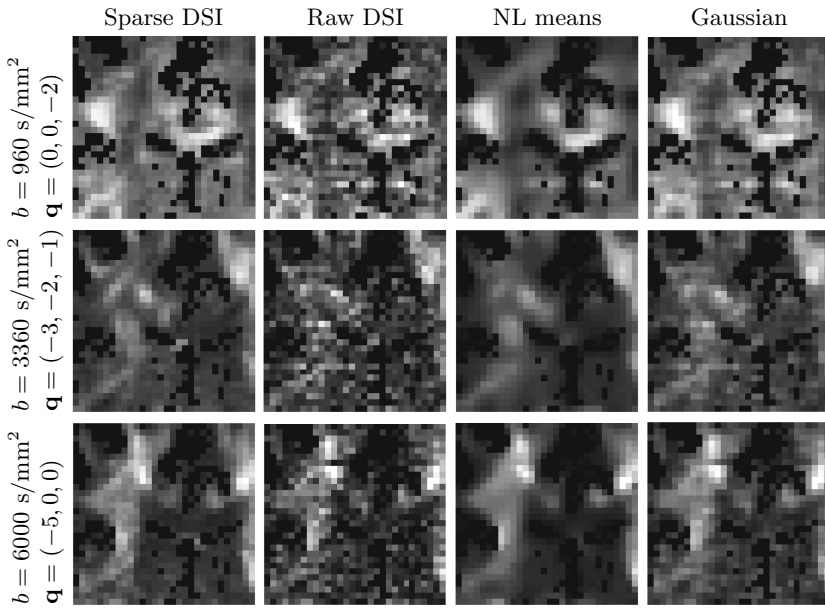| Methods	Gaussian $\sigma = 0.35$	NL means	Sparse DSI					
			$k = 100$	$k = 169$	$k = 225$	$k = 400$	$k = 900$	$k = 1600$
Subject 1	1.16	1.19	1.31	1.31	1.31	1.31	1.31	1.31
Subject 2	1.13	1.16	1.28	1.25	1.23	1.30	1.30	1.29

Fig. 3. Denoising the raw data DSI with our sparse DSI technique versus state-of-the-art non-local means (NLM) and Gaussian (optimal $\sigma = 0.35$) denoising.

Figure 5 shows the RMSE ratio between simple HALF DSI with symmetry and sparse DSI as a function of number of measurements. We also show the ODF field in Fig. 6 as a function of number of measurements. First, it is amazing to see that a learned DSI dictionary of a subject can be used to perform undersampled DSI on a different subject. It means that both dictionaries in Fig. 2 look similar and quantitatively yield comparable performances. Of course, as undersampling decreases, the overall field of ODF seems more noisy but the overall RMSE remains acceptable. At a total of 37, 29, and 21 measurements, we become worst than NL means, Gaussian smoothing and simple symmetry DSI in terms of RMSE. On the other hand, we observe that ODF profiles are degraded sooner as a function of undersampling. Note that the structured voxels with single fiber orientation in the CC, CT and SLF are well preserved all the way down to 29 measurements. However, although crossings are found for all undersampling, ODF peaks in crossing areas become less accurate below 58 measurements.
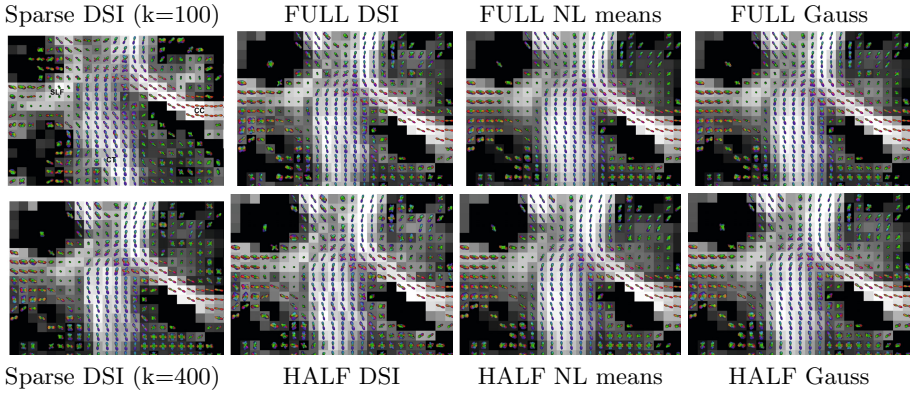
Sparse DSI (k=100)	FULL DSI	FULL NL means	FULL Gauss

Sparse DSI (k=400)	HALF DSI	HALF NL means	HALF Gauss

Fig. 4. Full ($d = 515$) DSI vs. Half DSI ($d = 258$) with respect to simple symmetry, Gaussian ($\sigma = 0.35$), NL means and sparse DSI denoising of subject 1 (k atoms).
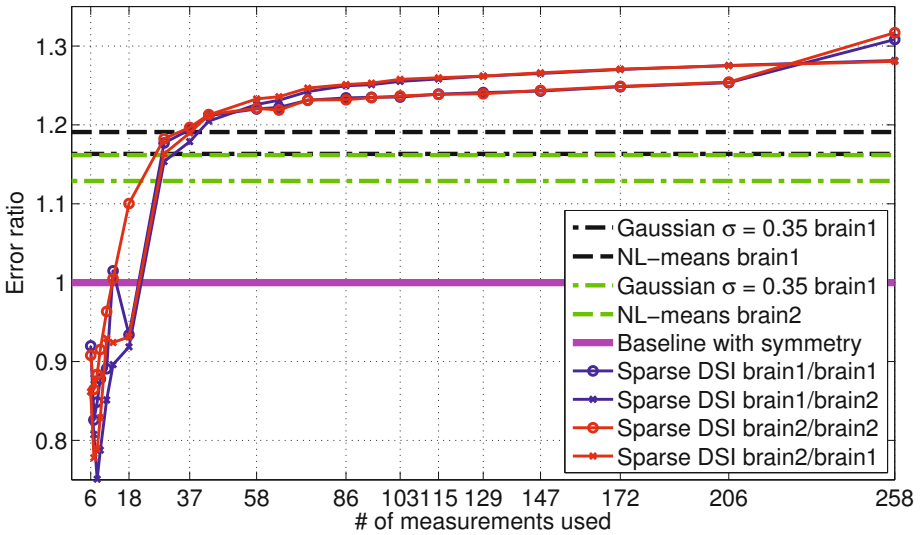

Fig. 5. Reconstruction error ratios for intra and inter-subject settings as a function of the number of measurements (k=100 atoms). brain1/brain1 is for the intra-subject case while brain1/brain2 is the inter-subject (Atoms learned on subject 1 and used the estimate the full DSI data of subject 2 using only a few measurements).
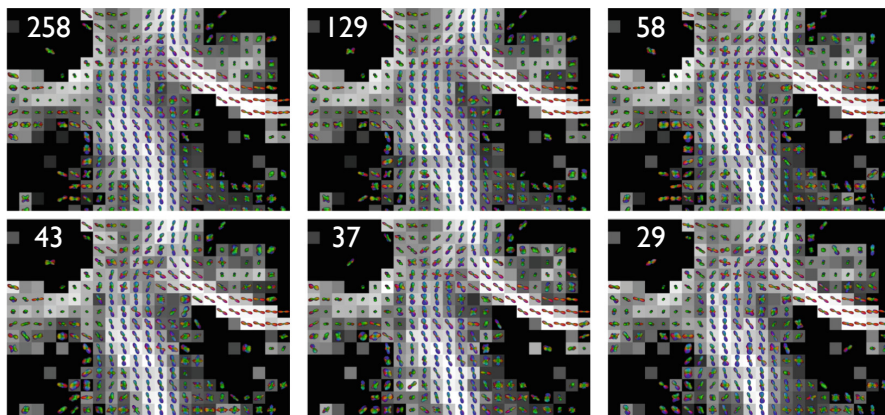
Fig. 6. Undersampled sparse DSI using learned dictionary from subject 1 to reconstruct DSI signal and diffusion ODFs of subject 2. In white is the number of measurements.

4   Discussion and Conclusion

Sparse coding applied to DSI data reveals the latent structure of the white matter. Sparse coding is however not compressed sensing (CS), as done for DSI in [7]. Although our technique attempts to infer full DSI data from undersampled acquisitions, there is no "sensing" in the technique. The idea is to learn the structure of raw DWI from a full DSI acquisition to either, denoise the DSI data, or use the learned dictionary of DSI profiles to perform undersampled DSI acquisitions and reconstructions. While [7,8,9] fix a priori the sparse representation of the data (*e.g.* spherical ridgelets) we propose here to estimate it. The technique proposed is attractive thanks to its little modeling assumptions and its limited number of parameters that can be estimated by cross-validation.

The key benefit of our method is its ability to perform denoising across all the DWI channels jointly, consequently enhancing the image quality in particular for noisy high b-values. While the technique of [10] uses DW images within a certain cone around the DW image being denoised, we propose to estimate the underlying structure from all directions and b-values. This is made possible by a proper whitening of the data in order to combine in the estimation multiple images corrupted by different noise levels.

Results have showed that with just half of the data (258 DWI), we can better predict the other 257 DWI than the classic symmetry procedure. This statement also holds even if we use as little as 40 q-space measurements. Our sparse DSI technique performs better than symmetrizing, Gaussian denoising or state-of-the-art NL means. Finally, beyond denoising, we have showed that learning the dictionary from one subject can be used to reconstruct full DSI dataset from an undersampled acquisition of a different subject. From now on, we could acquire around 40 measurements on new subjects and use the learned dictionaries to

reconstruct a full DSI data. Hence, we can have fast acquisitions to obtain high resolution DSI data. Therefore, DSI can be done at the price of HARDI!

Future work will be dedicated at optimizing the dictionary learning to enhance ODF reconstruction and also to find optimal sub-sampling strategies.

References

1. Wedeen, V.J., Hagmann, P., Tseng, W.Y.I., Reese, T.G., Weisskoff, R.M.: Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. Magnetic Resonance in Medicine 54(6), 1377–1386 (2005)
2. Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O.: Mapping the structural core of human cerebral cortex. PLoS Biology 6(7), e159 (2008)
3. Sijbers, J., den Dekker, A.J., Audekerke, J.V., Verhoye, M., Dyck, D.V.: Estimation of the noise in magnitude MR images. Magnetic Resonance Imaging 16(1), 87–90 (1998)
4. Koay, C.G., Özarslan, E., Basser, P.J.: A signal transformational framework for breaking the noise floor and its applications in MRI. Journal of Magnetic Resonance 197(2), 108–119 (2009)
5. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research 11, 19 (2010)
6. Coupé, P., Manjón, J.V., Gedamu, E., Arnold, D., Robles, M., Collins, D.L.: Robust rician noise estimation for MR images. Medical Image Analysis 14(4), 483–493 (2010)
7. Menzel, M.I., Tan, E.T., Khare, K., Sperl, J.I., King, K.F., Tao, X., Hardy, C.J., Marinelli, L.: Accelerated diffusion spectrum imaging in the human brain using compressed sensing. Magnetic Resonance in Medicine 66(5), 1226–1233 (2011)
8. Landman, B.A., Bogovic, J.A., Wan, H., ElShahaby, F.E.Z., Bazin, P.L., Prince, J.L.: Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. NeuroImage 59(3), 2175–2186 (2012)
9. Michailovich, O., Rathi, Y., Dolui, S.: Spatially regularized compressed sensing for high angular resolution diffusion imaging. IEEE Transactions on Medical Imaging 30(5), 1100–1115 (2011)
10. Tristán-Vega, A., Aja-Fernández, S.: DWI filtering using joint information for DTI and HARDI. Medical Image Analysis 14(2), 205–218 (2010)

Fiber Density Estimation by Tensor Divergence

Marco Reisert, Henrik Skibbe, and Valerij G. Kiselev

Medical Physics, University Medical Center Freiburg,
Breisacher Str. 60a, 79106 Freiburg, Germany

Abstract. Diffusion-sensitized magnetic resonance imaging provides information about the fibrous structure of the human brain. However, this information is not sufficient to reconstruct the underlying fiber network, because the nature of diffusion provides only conditional fiber densities. That is, it is possible to infer the percentage of bundles that pass a voxel with a certain direction, but the absolute number of fibers is inaccessible. In this work we propose a conservation equation for tensor fields that can infer this number up to a factor. Simulations on synthetic phantoms show that the approach is able to derive the densities correctly for various configurations. In-vivo results on 20 healthy volunteers are plausible and consistent, while a rigorous evaluation is difficult, because conclusive data from both MRI and histology remain elusive even on the most studied brain structures.

1   Introduction

The representation of neuronal fibers is of major interest in neuroscience and medicine. Diffusion weighted magnetic resonance imaging (DWI) has the potential to visualize non-invasively these fiber [1]. The quantitative reconstruction of the underlying fiber architecture is a great challenge, as DWI gives only indirect measures for the presence of neuronal fibers. One of the main problems is that the ordinary DWI measurement can only provide relative information. For each voxel one can infer a probability distribution of directions (fiber or diffusion orientation distributions), but these distributions do not give information about the total number of fibers within the voxel. There are attempts to measure fiber density directly, e.g. by determining the myelinization degree with bound pool fraction imaging [2] or with optimized diffusion measurements and multi-compartement models [3]. On the other hand, fiber tracking approaches can be used to infer tract-density maps [4,5], but the inferred maps are usually correlated closely with fractional anisotropy (FA) [6]. This work attempts to formulate a fiber conservation law, which is able to derive fiber density maps from ordinary (clinical suitable) DWI by employing the global coherence of the data. The idea is inspired by fluid dynamics, where mass conservation is expressed by a partial differential equation. A similar equation is formulated for tensor fields representing the underlying diffusion measurement. While in fluid dynamics the divergence equation can be derived from a strict integral formulation, our derivation remains 'empirical', but simulations show that the obtained solutions work well. The proposed equation is solved via a finite element method on a standard PC in a few minutes.

2   Method

In fluid dynamics, the mass conservation is expressed by the equation $\partial_i(v_i(\boldsymbol{r})\rho(\boldsymbol{r})) = 0$, where $\boldsymbol{v} = (v_1, v_2, v_3)$ and ρ are the velocity field and the particle density at a point $\boldsymbol{r} = (r_1, r_2, r_3)$, respectively, and $\partial_i = \partial/\partial r_i$. (We use Einstein's summation convention throughout this paper, all indices appearing more than once are implicitly summed). In fluid flows each point $\boldsymbol{r}$ is assigned a unique velocity $\boldsymbol{v}(\boldsymbol{r})$. The case of fibers is different. At a finite resolution it is possible that several fibers with different directions cross a volume element. Detailed information about the fiber orientations is thus provided by the absolute joint fiber distribution $p(\boldsymbol{n}, \boldsymbol{r})$ that depends on the unit vector $\boldsymbol{n} \in S_2$ defining the fiber direction and position $\boldsymbol{r}$. This function is not directly accessible with DWI. The DWI observable quantity is rather the conditional distribution of directions $p(\boldsymbol{n}|\boldsymbol{r})$, which does not provide any information about the absolute density of fibers $\rho(\boldsymbol{r})$. These quantities are related as $p(\boldsymbol{n}, \boldsymbol{r}) = p(\boldsymbol{n}|\boldsymbol{r})\rho(\boldsymbol{r})$ provided the normalization $\int_{S_2} p(\boldsymbol{n}|\boldsymbol{r}) \, d\boldsymbol{n} = 1$. So, our goal is to is to find a 'fiber density' $\rho(\boldsymbol{r})$ for a given conditional orientation density $p(\boldsymbol{n}|\boldsymbol{r})$. However note, that we can infer ρ only up to a global normalization factor, that is, we cannot infer an 'absolute' fiber density map. This problem is ill-posed, which can be illustrated with an example of isotropic $p(\boldsymbol{n}|\boldsymbol{r})$. Such a distribution can result from fibers with any density. We thus need some additional prior knowledge to regularize the problem. We seek for a cost functional that favors reasonable configurations. The cost functional is constructed by using the principle of fiber continuity [7], which assumes that fibers appear to be locally straight. The main idea is to require $n_i \partial_i p(\boldsymbol{n}, \boldsymbol{r})$ to be small in magnitude. The term $n_i \partial_i p(\boldsymbol{n}, \boldsymbol{r})$ is the change in the number of fibers with direction $\boldsymbol{n}$ when moving in direction $\boldsymbol{n}$. However, this assumption is too restrictive not being able to cope with curved fibers. Thus, we want to build costs that are more selective. Therefore, consider the vector field $n_i n_j \partial_j p(\boldsymbol{n}, \boldsymbol{r})$. This vector points in an arbitrary selected direction $\boldsymbol{n}$ and its magnitude equals the gradient of the fiber density in this direction. We sum this vector over all fiber directions to obtain an integral characteristics of a voxel:

$$k_i(\boldsymbol{r}) = \int_{S_2} n_i n_j \partial_j p(\boldsymbol{n}, \boldsymbol{r}) \, d\boldsymbol{n} = \partial_j \int_{S_2} n_i n_j \, p(\boldsymbol{n}, \boldsymbol{r}) \, d\boldsymbol{n} = \partial_j P_{ij}(\boldsymbol{r}) \qquad (1)$$

Note that this vector field $\boldsymbol{k} = (k_1, k_2, k_3)$ only depends on the projection $P_{ij}(\boldsymbol{r}) = \int n_i n_j p(\boldsymbol{n}, \boldsymbol{r})$, which is an ordinary Cartesian rank-2 tensor. As we know that $p(\boldsymbol{n}, \boldsymbol{r}) = p(\boldsymbol{n}|\boldsymbol{r})\rho(\boldsymbol{r})$, the tensor P_{ij} can also be expressed in terms of the projection of the conditional $T_{ij}(\boldsymbol{r}) = \int n_i n_j p(\boldsymbol{n}|\boldsymbol{r})$ as follows:

$$P_{ij}(\boldsymbol{r}) = \rho(\boldsymbol{r}) \int_{S_2} n_i n_j p(\boldsymbol{n}|\boldsymbol{r}) \, d\boldsymbol{n} = \rho(\boldsymbol{r}) \, T_{ij}(\boldsymbol{r}). \qquad (2)$$

Considering the example of parallel fibers, it is clear that $\boldsymbol{k} = 0$ expresses the fact that the fibers do not terminate. So, a non-zero $\boldsymbol{k}$ has to be suppressed in this case. On the other hand, $\boldsymbol{k}$ defines the curvature field for bending fibers[1] and should not be penalized

---

[1] For simplicity we give a 2D example. The bending configuration is given by $(n_1^{\text{fib}}, n_2^{\text{fib}}) = (-r_2, r_1)/|\boldsymbol{r}|$ from which we have $(P_{ij}) = \frac{1}{|\boldsymbol{r}|^2} \begin{pmatrix} r_2^2 & -r_1 r_2 \\ -r_1 r_2 & r_1^2 \end{pmatrix}$. Computing the partial derivatives we get $k_j = \partial_i P_{ij} = -(r_1, r_2)/|\boldsymbol{r}|^2$ and hence $|\boldsymbol{k}| = 1/|\boldsymbol{r}|$ which was to show.

in that case. To implement such a selected suppression of $\boldsymbol{k}$, we use the fact that the curvature vector of a fiber is orthogonal to its tangent. Instead of looking for ρ's that minimize the squared norm of the field $\boldsymbol{k}$, we use the tensor T_{ij} as a metric tensor to project out the pure curvature contributions. That is, we look for a fiber density that minimizes the cost functional

$$J_{\text{TD}}(\rho) = \int_{\Omega} k_i T_{ij} k_j \, d\boldsymbol{r} = \int_{\Omega} (\partial_a P_{ia}) T_{ij} (\partial_b P_{jb}) d\boldsymbol{r} \tag{3}$$

We used T_{ij} instead of P_{ij} as metric tensor, because this keeps our objective quadratic in ρ, and hence, we can solve the problem by inverting a linear system. Note that for a constant isotropic field $T_{ij} = \delta_{ij}$ our equation reduces to the integral over the gradient magnitude of ρ which is equivalent to Laplace's equation. Thus, more isotropic tensors T_{ij} lead to more spatially homogeneous fiber densities.

2.1   Boundary Conditions and Implementation via Finite Elements

As the objective as given in equation (3) is of differential nature we have to keep care what happens at the boundary $\partial\Omega$ of the white matter domain. The weakest assumption is a Neumann boundary condition, that is $v_j \partial_j \rho(\boldsymbol{r}) = 0$ for $\boldsymbol{r} \in \partial\Omega$ and $\boldsymbol{v}$ orthogonal to the boundary. However, in this setting our minimization problem is solved by the trivial solution $\rho = 0$, hence we need additional assumptions to keep the solution away from being zero. The simplest way is to set ρ on the boundary (in the transition area of white and gray matter) to a specific value (Dirichlet condition). Unfortunately, we need an idea of the amount of fiber terminals to set ρ to reasonable values. Of course, a uniform distribution is an option. Another way is to modify the objective. We figured out two ways: one can subtract the total number of fibers $J(\rho) = J_{\text{TD}}(\rho) - \int_{\Omega} \rho$ such that the consistency term J_{TD} is minimized and simultaneously the total number of fibers $\int_{\Omega} \rho$ is maximized. Secondly, one can add a quadratic deviation from an initial guess ρ_0, that is $J(\rho) = J_{\text{TD}}(\rho) + \lambda \int_{\Omega} |\rho - \rho_0|^2$. If $\rho_0 = 1$ and λ sufficiently small, this approach converges towards the first option. We will discuss the different options in the experiments.

   To solve the minimization problem we followed a finite element (FEM) discretization which is ideally suited for our problem. We cannot give here a complete introduction to FEM and refer to [8], but shortly propose which element we used and what is important during discretization. To stay as close as possible to the initial Cartesian discretization of the data we use a trilinear finite element, that is, the elements are cubes with corners in the center of each voxel, where the interpolation function within this cube is the ordinary trilinear one frequently used in computer graphics. Thus, the unknown variables are just the density values at the voxel positions. The advantage is that we can treat the tensor field $T_{ij}(\boldsymbol{r})$ in the same manner. The values of T_{ij} are known at the voxel centers and the value within the finite element can also be computed by trilinear interpolation. Each unknown variable is associated with a basis function. The density ρ is a linear combination of all basis functions $\rho(\boldsymbol{r}) = \sum_{\beta} a_{\beta} \phi_{\beta}(\boldsymbol{r})$ where β runs over all voxels, a_{β} are the unknowns and ϕ_{β} are the basis functions. In FEM we have to compute for all pairs of basis function ϕ_{μ}, ϕ_{β} the integral $M_{\mu,\beta} = \int_{\Omega} (\partial_a T_{ia} \phi_{\mu}) T_{ij} (\partial_b T_{jb} \phi_{\beta}) d\boldsymbol{r}$, which forms the matrix which one has

to invert[2]. For the computation of the integrals we used the MATLAB symbolic tool-box. Assembling the matrix with a standard desktop PC needs about 3 minutes for half a million unknowns. Solving the problem by an iterative method is in the same time range (we used the MATLAB *minres* algorithm with a tolerance of 10^{-5}).

3   Experiments

We start with simulations on synthetic data to show that the solutions of our equations give quantitative results. Three configurations are considered: a fanning, a bending and a crossing. The fanning was generated by parabolas, where the opening degree varies with the y-axis section[3]. For the crossing we varied the angle and the relative density of the tracts. To generate the tensor field T_{ij} we followed the approach described above, i.e. $T_{ij} = n_i^{\text{fib}} n_j^{\text{fib}}$. Of course, the field is always normalized such that $T_{ii} = 1$. The configurations were rendered on a $128 \times 128 \times 3$ matrix. Finally, we smoothed the field T_{ij} with a Gaussian of width 1 voxel. We investigated the three different boundary as-sumption discussed above: using Dirichlet boundaries in the regions where the fibers terminate (DIR), or modifying the objective by subtracting the total number of fibers (LIN), or adding the quadratic deviation from a uniform density $\rho_0 = 1$ (QUAD). In Figure 1 we show densities inferred with DIR/LIN/QUAD assumptions. The Dirich-let (DIR) boundaries are the 'easiest' case, as the most prior knowledge is put in. All three configurations are well resolved in this case. For the fanning we checked the fiber number conservation by integrating the density along the y-axis. Ideally, it should be a constant value. And, in fact, up to some fluctuation due to discretization it is a con-stant. During the narrowing the packing is getting dense, and during the widening the density decreases. The two crossing tracts (crossing angle of 60 degrees) are assumed to have a density ratio 2:1. For an orthogonal crossing this is expressed by the fact that the eigenvalues of T_{ij} in the crossing region have also the ratio of 2:1. Also the DIR boundary values were set to this ratio. Obviously the crossing is nicely resolved, also quantitatively the values within the crossing regions are the sum of the values of the two inflowing tracts. Finally, the result for the bending is, as desired, a constant density. For the LIN approach the results are a bit different. For the crossing the results do not change, which is encouraging. Without putting any prior knowledge about the in- and out-flowing fibers the equation is able to infer the relative densities perfectly. For the bending case (Figure 1f),g)) we get different results, tracts with high curvature are pe-nalized, although we have shown above that for a perfect bending $T_{ij}\partial_j P_{aj} = 0$. That is, curved fibers should not be penalized. We found that this is basically a discretization artifact. The sharper the bending the higher the errors in the discretized computation of $T_{ij}\partial_j P_{aj}$. However, it is not astonishing that the solution is so sensitive. Theoretically,

---

[2] For the computation of $M_{\mu,\beta}$ there is an important note. To compute the product $T_{jb}\phi_\beta$ within an element one has to first multiply the values of T_{jb} and ϕ_β at the element corners and then in-terpolate within the element. We found that it is not possible to interpolate both independently and then multiply the interpolants.

[3] The fanning is represented by a family of parabolas $y_a(x) = a(1 + bx^2)$ where $a \in (-d/2, d/2)$ and d is the thickness of the bundle. The parameter b controls the degree of fanning
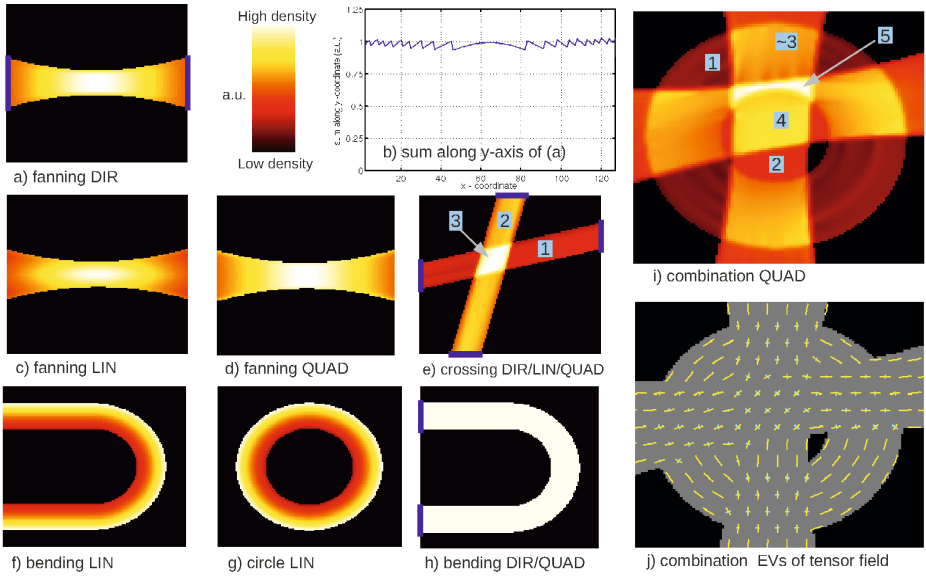
Fig. 1. Fiber densities of the synthetic phantoms. The blue bars in a,e,h) indicate the Dirichlet boundaries. **a)** The fanning with DIR boundary. **b)** The sum along the y-axis of the densities in (a) to show the fiber number preservation. **c)** Fanning with LIN approach: the central, non-curved fibers are emphasized. **d)** For the QUAD approach there is nearly no difference to DIR. **e)** The crossing is well resolved for all three boundary approaches, the density values in the indicated regions have the desired ratio 1:2:3. **f,g)** The LIN approach tends to give lower density to curved fibers, which is a discretization artifact. **h)** for the DIR and QUAD approach the density is constant over the bending. **i)** density of complex configuration with annotated ratios. **j)** the eigenvectors of the tensor T_{ij} representing the experimental data.

there is huge family of solutions to the LIN based equation. Any density profile is a valid solution as long as the profile do not change along the bundle. The situation is similar like in fluid dynamics: imagine a non-viscous compressible fluid flowing in a tube with constant velocity $\boldsymbol{v} = (v_0, 0, 0)$. Any density profile $\rho(x, y, z) = \rho_0(y, z)$ is a solution to the equation $\partial_i(v_i \rho) = 0$. The QUAD approach resolves this ambiguity, which can be seen in Figure 1h). Actually, the QUAD approach can be seen as a small Tikhonov-regularization with $\lambda \int \rho^2$ to make the equation less sensitive to the discretization artifact, because $\int |\rho - 1|^2 = \int \rho^2 - 2 \int \rho + \text{const}$. We found a regularization parameter of $\lambda = 1/n$ to work well, where n is the number of unknown variables. The discretization artefacts disappear and for all configurations (Figure 1d,e,h) the densities stay quantitative, i.e. the densities of the crossing keep the assumed ratios, for the fanning the fiber number stays constant and for the bending we get a uniform density. In a further experiment (Figure 1i,j) we found that the results also generalize: a combination of a fanning, crossing and bending. Each of the crossing tracts has a density of 2 at its most narrow location, while the 'circle' configuration has density 1. Up to some small discretization artifacts the densities are well inferred. We want to emphasize that although our approach soley relies on a tensor representation, the densities for the threefold crossing are predicted well.

3.1   In-Vivo Data

To investigate our approach on real DTI data we considered 20 scans of healthy volunteers at a b-value of $1ms/\mu m^2$ with 61 diffusion directions and an isotropic resolution of $2mm^3$. A white matter probability map was generated with SPM (Version 5, *http://www.fil.ion.ucl.ac.uk/spm/*) on a T1-weighted scan, which was co-registered to the b_0-scan of the diffusion sequence. The probability maps are thresholded at 0.1 to get the simulation domain Ω. The diffusion weighted images were up-sampled to a $1mm^3$-resolution. We decided to base the conditional distribution $p(\boldsymbol{n}|\boldsymbol{x})$ purely on the diffusion tensor, thus the following experiments are also appropriate in a clinical setting with low diffusion weighting. We used the formula $p(\boldsymbol{n}|\boldsymbol{x}) \propto e^{-n_i D_{i,j}^{-1} n_j r_0^2/(2t)}$, where D^{-1} is the inverted diffusion tensor, r_0 a distance and t the diffusion time. That is, $p(\boldsymbol{n}|\boldsymbol{x})$ is the probability that a particle is displaced by an amount of r_0 in direction $\boldsymbol{n}$ within the time t. It is important to control the sharpness (entropy) of the distribution. A high entropy will lead to more uniform fiber densities than a low entropy. The parameter r_0^2/t controls this sharpness. We found a value of $r_0^2/t = 25\mu m^2/ms$ appropriate. Of course, a more detailed investigation is needed at this point. The tensor T_{ij} is computed via $T_{ij}(\boldsymbol{r}) = \int n_i n_j p(\boldsymbol{n}|\boldsymbol{r})$ numerically and normalized such that $T_{ii} = 1$. To get a non-vanishing solution the above introduced QUAD approach was used.

In Figure 2a-f) the fiber density found by our approach (FD) is compared to fractional anisotropy (FA) and tract densities (TD) computed by a global fiber tracking approach [5]. While TD is closely correlated to FA the proposed densities are different and do not show the expected [9] picture: FD does not have the maximal dense regions in the Corpus Callosum (CC) and along the cortical spinal tract. The densest regions are around crossing areas and the superior longitudinal fascicle. However note, that the angular maximum map $ma(\boldsymbol{r}) = \max_{\boldsymbol{n}} p(\boldsymbol{n}, \boldsymbol{r}) = \rho(\boldsymbol{r}) \max_{\boldsymbol{n}} p(\boldsymbol{n}|\boldsymbol{r})$ (not shown) resembles much more closely TD or FA, respectively. To get a better impression of the distribution over the whole brain Figure 2g-j) shows cortical projections of FD. At each point of the white matter isosurface the local average (a neighborhood of 5 voxels) of the fiber density is displayed. While the frontal and parietal lobe show an increased FD, the temporal and occipital lobes appear to be less dense. To get a quantitative picture over the whole group of subjects the WFU pickatlas (*http://fmri.wfubmc.edu/software/PickAtlas*) was employed to segment in Talairach Daemon Level 2 (lobes) regions. A statistical analysis over the mean FD in the regions is shown in Figure 3,left). For comparison we normalized FD such that the total fiber number $\int \rho$ is the same for all subjects. The density for the different regions is stable over the group and differences between the regions are significant. It is apparent that frontal/parietal regions are more dense than the occipital/temporal regions. Figure 3 shows histograms over the whole white matter area for all 20 subjects. Finally, in Figure 3 (bottom) the CCs of all 20 subjects are shown. The Splenium and Genu appear to have the highest densities while the region around the Isthmus the lowest. There is a small trend of more dense packings towards the anterior part of the CC body, which seems to be consistent with bound pool imaging [2] and axon density indices [3].
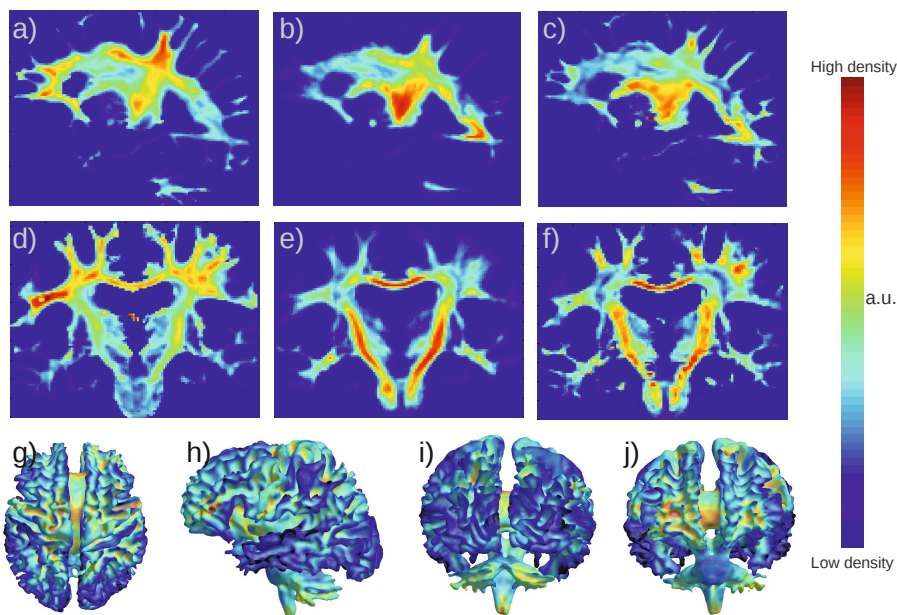
Fig. 2. a,d) A coronal (at the level of the mid-body of the Corpus Callosum) and saggital section of the fiber densities obtained from the proposed approach. **b,e)**: Tract-densities inferred by global fiber tracking. **c,f)**: Fractional Anisotropy. **g-j)** Cortical projections of the fiber densities obtained from the proposed approach. The temporal and occipital regions appear to be less populated.
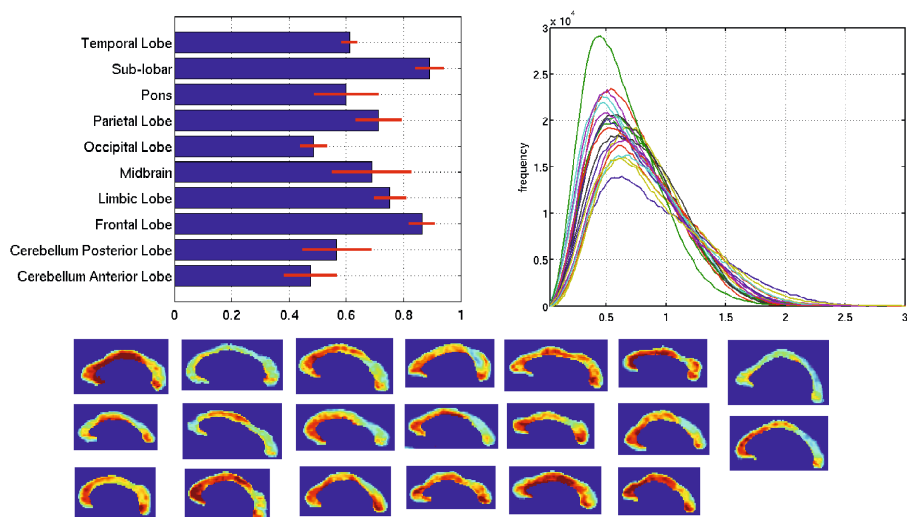

Fig. 3. Top: Statistics over 20 healthy subjects. **Top left**: mean densities in different regions with group-deviations. **Top right**: Distribution of densitites over the whole brain of all 20 subjects. **Bottom**: Fiber densities of the Corpus Callosum for 20 healthy subjects with Genu on the left and Splenium on the right.

4   Conclusion

The main contribution of this work is the proposal of a PDE for tensor-fields, which reflects the fiber conservation law: fibers are not allowed to end somewhere in the white matter. The PDE was implemented with a FEM approach and is solved in a few minutes on a standard PC. Simulations on synthetic data show that the results are quantitative even in the absence of boundary conditions. Although our approach relies on tensor information only, we were able to show that phantoms with a threefold crossing can be resolved. On real DTI data the derived fiber densities appear partially to be consistent with other findings [3,2], while a rigorous evaluation with histological findings [9] remains challenging.

References

1. Jones, D.K. (ed.): Diffusion MRI: Theory, Methods and Applications. Oxford University Press (2010)
2. Stikov, N., Perry, L.M., Mezer, A., Rykhlevskaia, E., Wandell, B.A., Pauly, J.M., Dougherty, R.F.: Bound pool fractions complement diffusion measures to describe white matter micro and macrostructure. Neuroimage 54, 1112–1121 (2011)
3. Alexander, D.C., Hubbard, P.L., Hall, M.G., Moore, E.A., Ptito, M., Parker, G.J., Dyrby, T.B.: Orientationally invariant indices of axon diameter and density from diffusion mri. NeuroImage 52, 1374–1389 (2010)
4. Calamante, F., Tournier, J., Jackson, G., Connelly, A.: Track-density imaging (tdi): super-resolution white matter imaging using whole-brain track-density mapping. Neuroimage 53 (2010)
5. Reisert, M., Mader, I., Anastasopoulos, C., Weigel, M., Schnell, S., Kiselev, V.: Global fiber reconstruction becomes practical. Neuroimage 54, 955–962 (2011)
6. Roberts, T., Liu, F., Kassner, A., Mori, S., Guha, A.: AJNR Am. J. Neuroradiol. 26, 2183–2186 (2005)
7. Reisert, M., Kiselev, V.: Fiber continuity: An anisotropic prior for odf estimation. IEEE Trans. Med. Imaging 30, 1274–1283 (2011)
8. Chung, T.: Computational Fluid Dynamics. Cambridge University Press (2002)
9. Burgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J., Zilles, K.: White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. Neuroimage 29, 1092–1105 (2006)

Estimation of Extracellular Volume
from Regularized Multi-shell Diffusion MRI[⋆]

Ofer Pasternak[1], Martha E. Shenton[1,2], and Carl-Fredrik Westin[1]

[1] Brigham and Women's Hospital, Harvard Medical School, Boston MA
[2] VA Boston Healthcare System, Brockton, MA

Abstract. Diffusion MRI measures micron scale displacement of water
molecules, providing unique insight into microstructural tissue architec-
ture. However, current practical image resolution is in the millimeter
scale, and thus diffusivities from many tissue compartments are averaged
in each voxel, reducing the sensitivity and specificity of the measurement
to subtle pathologies. Recent studies have pointed out that eliminating
the contribution of extracellular water increases the sensitivity of the
diffusion measures to tissue architecture. Moreover, in brain imaging,
estimation of the extracellular volume appears to indicate pathological
processes such as atrophy, edema and neuroinflammation. Here we study
the free-water method, which assumes a bi-tensor model. We add low b-
value shells to a regular DTI acquisition and present methods to improve
the estimation of the model parameters using the extra information. In
addition, we define a Laplace-Beltrami regularization operator that fur-
ther stabilizes the multi-shell estimation.

1 Introduction

Diffusion MRI (dMRI) measures the displacement of water molecules, which in
a typical brain imaging experiment displace a few tens of microns. This makes
dMRI sensitive to normal and pathological architecture in the cellular scale. In-
deed, dMRI based sequences, such as diffusion tensor imaging (DTI), have been
found extremely useful in identifying subtle changes, especially in white-matter,
that occur in normal development, as well as in many types of diseases, abnormal-
ities and disorders [1]. Nevertheless, current image resolution is in the millimeter
scale, introducing partial volume of different tissue types - white matter, gray
matter, glia cells, cerebrospinal fluid (CSF) - which reduces the sensitivity and
specificity of most indices derived from dMRI and DTI [2].

Correcting for extracellular water is required to eliminate CSF contamina-
tion, thus improving DTI's sensitivity in the vicinity of the ventricles [3] and
important for the delineation of fibers such as the fornix [4,5]. Moreover, the
fractional volume of the extracellular water, relative to the remaining hindered
or restricted water molecules, appears to provide important information with

regard to pathological processes that modify the interstitial extracellular space, such as edema [4], neuroinflammation [6] and atrophy [7]. Indeed, the extracellular volume was shown to be sensitive to pathologies that appear in aging [7], schizophrenia [8], Multiple-Sclerosis [6], and Alzheimer's disease [9].

In this work we present a method to estimate the extracellular volume and the diffusivities of the remaining tissue using a multi-shell dMRI acquisition. Motivated by increasingly more designs that already choose multi-shell for clinical studies, we enable estimation of free-water from such acquisitions in a novel way that utilizes the additional information instead of discarding it.

2   The Free-Water Model

The free-water model estimates and corrects for the contribution of the extracellular space [4]. The model assumes that the diffusion signal originates from two molecular compartments, co-existing within a voxel, with slow exchange between the compartments [4]:

$$A_i(D, f) = f \exp(-b_i q_i^T D q_i) + (1 - f) \exp(-b_i d) . \tag{1}$$

Here, A_i is the signal (normalized by the b_0) of the ith applied diffusion gradient with orientation q_i, and b-value b_i. The first term reflects the tissue compartment, where D is the diffusion tensor of this compartment and f is the relative contribution of the compartment. The second term reflects an isotropic compartment, with a fixed diffusion coefficient, d, set to the diffusion coefficient of water in body temperature, $3 \times 10^{-3} mm^2/s$. Thus, the isotropic compartment models free-water, i.e., molecules that do not experience hindrance. In a cellular environment such as brain tissue, free-water can only be in the extracellular space.

The free-water model adds only one more parameter, f, to the DTI model. However, unlike DTI, the fitting of this bi-exponential model is highly unstable [10]. Pierpaoli et al., proposed to stabilize the fit by measuring multiple b-shells that achieve high b-values, requiring a large number of measurements [11]. However, this approach required lengthy scans, and moreover, at high b-values the tissue compartment is no longer adequately described by a single exponent [10]. Pasternak et al., introduced a regularization framework that allowed the estimation of the free-water model from a single-shell DTI acquisition [4]. However, this method requires smoothing, which may reduce the sensitivity to subtle details or pathologies.

In our work we combine principles from these two approaches. Similar to [11], we acquire multi-shell data. However, our approach only requires lower b-valued shells. We present a novel way of estimating the free-water model parameters by separating the higher shells from the lower shells. We then use these estimates to initialize a spatial tensor regularization refinement step similar to [4], to result with the final estimates.

3   Multi-shell Free-Water Estimation

Our approach has three steps. In the first step (Sec. 3.1) the diffusion tensor of the tissue compartment is estimated from the higher b-valued shells. In the second step (Sec. 3.2) the estimated tensors along with the lower b-valued shells are used to estimate the extracellular volume. In the final step (Sec. 3.3) we use these estimates as an initial guess to a regularized minimization process that takes in account the entire scope of the data.

3.1   Estimation of Free-Water Eliminated Tensors

The diffusivities of white and gray matter are considerably lower than those of free-water or CSF. Typically, in single shell DTI with a b-value around 1000 s/mm^2, healthy brain tissue has a quite homogeneous mean diffusivity of around $0.8mm^2/sec$, 3-4 times slower than free-water. Therefore, the free-water signal is expected to decay faster than tissue, e.g., with a b-value of $900mm^2/sec$ the tissue decays to 49% of the signal while free-water decays to 7% of the signal. Figure (1) shows an example of a multi-shell acquisition for a range of b-values. This range is achieved by modifying the diffusion gradient amplitude for fixed diffusion times. As expected, the free-water signal (mainly seen in the ventricles and around the parenchyma) attenuates faster than other brain tissue. The free-water signal diminishes completely into the noise floor for the higher b-values.

Also dependent on the b-value is the apparent diffusion coefficient of a partially volumed voxel, which can be expressed as [5]:

$$D' = b_i^{-1} \log(f \exp(-b_i q_i^T D q_i) + (1 - f) \exp(-b_i d)) . \tag{2}$$

This means that D' is not the average of the diffusivities of the two compartments, but is closer to D as b increases. Therefore, we suggest that a tensor estimated only from the higher b-valued shells (at least two shells are required), will minimize the contribution of the free-water and will be a good estimation for the tissue compartment tensor D. However, these shells cannot have a too-high b-value, for two reasons: (1) Sensitivity is reduced to tissue with high diffusivities, such as along axons, where the signal has already diminished in the higher
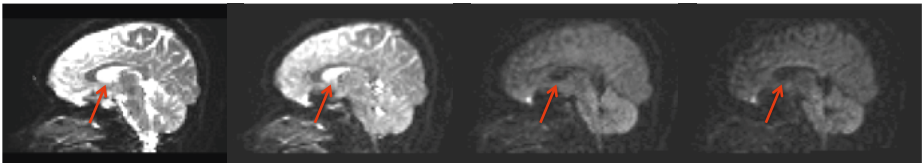

Fig. 1. Multi-shell acquisition. Diffusion signal for increasing b-values of (left to right) 0, 200, 900 and 1400 mm^2/sec in a mid-sagittal plane. The signal from free-water, such as in the ventricles (red arrow), decays into the noise floor faster than the signal of brain tissue.

b-values; (2) Since cellular tissue is heterogenous, the assumption of an exponential decay no longer holds for higher b-values [10], potentially biasing the estimated tensor.

We calculate D_H, the apparent diffusion tensor for the high b-valued shells, as an estimator for D by minimizing:

$$\sum_{i \in G_H} ||\tilde{b_0} \exp(-b_i q_i^T D_H q_i) - \hat{E}(q_i)|| \; , \tag{3}$$

where G_H are the indexes of all the applied gradients within the high b-valued shells, and $\hat{E}(q_i)$ is the signal (not normalized by the acquired b_0) of these images. The acquired b_0 reflects the contribution of all spins within the voxel, including from free-water. Therefore, the estimation of D_H requires the estimation of $\tilde{b_0}$, which is the baseline image that would have been acquired in the case that the tensor D_H was the only component in the voxel. We minimize equation (3) using a linear least square (LLS) approach with $ln(\tilde{b_0})$ as one of the free parameters [3].

3.2 Estimation of Extracellular Volume

We estimate f, which reflects the extracellular relative volume in a voxel, using the low b-valued shells, which are in the range that still has signal from free-water. Given D_H as an estimate for D, we can calculate f_L as an estimate for f using LLS by defining:

$$f_L = (y^T y)^{-1} y^T x \; , \tag{4}$$

where $x = A_i/A_0 - \exp(-b_i d)$ and $y = \exp(-b_i q_i^T D_H q_i) - \exp(-b_i d)$, and $i \in G_L$ being the indexes of the applied gradients in the low b-valued shells. A_i are the attenuation images defined in Eq. (1). Unlike our approach here, the single-shell free-water map estimation is initialized by the b_0 image alone, normalized by baseline values that assumed knowledge of voxels that have no tissue, and voxels that have no free-water [4]. This implicitly assumes that the T2 weighted images behave similarly across the entire brain, and that there are such baseline voxels. These assumptions are no longer required if using f_L and D_H as initialization.

3.3 Regularization of the Fitting

The initialization of the extracellular volume, f_L, and the tensor of the tissue compartment, D_H, were designed to be close to the real values, D and f, but could be biased due to either residuals of the free-water signal that remains in the higher b-valued shells, or due to high diffusivity within the tissue, the signal of which could disappear when using the higher b-valued shells. To avoid this potential bias we introduce a third step in which the entire information from all shells is fitted to the model. The fitting is initialized with the estimates provided in the previous two steps, and is stabilized by regularization. We use a regularization method based on the one proposed in [4] by minimizing the following functional:

$$L(D, f) = \int_\Omega \sum_{i \in G} ||A_i - \hat{A}_i|| + \alpha \sqrt{|\gamma(D)|} \; . \tag{5}$$

Here, Ω includes all voxels of interest, G are the indexes of all applied gradients and A_i are their signal normalized by the b_0. The parameter α scales the contribution of the Polyakov action regularization term (typically $\alpha = 1$ [4]), with $|\gamma(D)|$ as the determinant of the induced metric. Using the Einstein summation convention this metric has the form $\gamma_{\mu\nu}(x) = \partial_\mu X^i \partial_\nu X^j h_{ij}(X)$. Differing from the implementation in [4] and following the findings in [12] we do not use an affine-invariant metric to describe distances between tensors but rather use a Euclidean metric and the canonical tensor representation. In a regularization scheme the effect of metric selection is minimal [12], yet here the Euclidean metric simplifies the calculations considerably. As a result the vector X has the elements $X = [D_{xx}, D_{yy}, D_{zz}, 2D_{xy}, \sqrt{2}D_{yz}, \sqrt{2}D_{xz}, x, y, z]$. The matrix $H = \{h_{ij}\}$ is the spatial-feature metric that for a Euclidean tensor space is simply a 9×9 diagonal matrix, with 1 for the last three diagonal entries (the spatial domain) and a constant for the remaining 6 diagonal entries. This leads to the motion equations for the six unique tensor elements, D^j with $j \in \{1, 2, ..., 6\}$:

$$D_t^j = \sum_{i \in G} b_i(A_i - \hat{A}_i) \exp(-b_i q_i^T D q_i)\left(q_i^T \frac{\partial D}{\partial X_j} q_i\right) + \frac{\alpha}{\sqrt{|\gamma|}} \partial_\mu \sqrt{|\gamma|}(\gamma^{\mu\nu}\partial_\nu X^j) \tag{6}$$

and for the fractional volume parameter:

$$f_t = \sum_{i \in G} -b_i(A_i - \hat{A}_i)\left(\exp(-b_i q_i^T D_t q_i) - \exp(-b_i d)\right) . \tag{7}$$

Importantly, due to the use of the Euclidean metric, and unlike the motion equations derived in [4], equation (6) does not have any Christoffel numbers, and therefore its calculation is simpler and faster. The second term in equation (6) is the Laplace-Beltrami operator, which is a piece-wise smooth, edge preserving tensor regularization operator [4]. The final result is thus the parameters f and D that best fit the data while maintaining continuous tissue representation.

4   Experiments and Results

Multi-shell Acquisition Schemes. We test the multi-shell estimation on two types of data sets that were obtained with multi-shells. The first data uses an acquisition optimized for the free-water estimation, having a single b=0, 3 × b=50, 6 × b=200, 10 × b=500, 30 × b=900 and 16 × b=1400, with gradient orientations designed as nested platonic solids, which means that each shell is rotationally invariant, and the shells complement each other to a rotationally invariant scheme. Data was acquired on a 1.5T scanner with 2.5mm isotropic voxels and takes 9:20 minutes. We use the b=1400 and b=900 shells to estimate D_H, and the remaining shells to estimate f_L. The second data set demonstrates adding lower shells to an existing DTI gradient scheme with b=900 and the scanner default 64 gradient orientations. We added a shell at b=400 with 10 measurements and a shell at b=100 with 6 measurements, adding 2:40 minutes to the acquisition time. We use the b=900 and b=400 shells to estimate D_H.

We use b=100 and b=400 to estimate f_L. This data was acquired on a 3T scanner with 2mm isotropic voxels.

Implementation. All data was motion and eddy-current corrected. We used 3D-Slicer's tensor estimation to calculate D_H by first omitting all images in G_L, including the b_0. We used Matlab (Natick, MA) to calculate f_L. The optimization was done using a Matlab code that was provided in [4]. The code was changed to accept multiple b-values, to regularize using a Euclidean metric, and to use D_H and f_L as initialization. The complete analysis for a whole brain takes less than 15 minutes on a 64-bit Linux machine with Xeon-E5530 processors, without taking advantage of multiple cores.

Optimized Multi-Shell Acquisition. Both the multi-shell (Fig. 2[A]) and single-shell (using the b=900 shell ; Fig. 2[B]) estimations provide similar free-water maps, showing high values in the ventricles, and low values in the brain tissue, nicely depicting the extracellular volume. The color by orientation maps are similar as well. To better evaluate the differences between the maps, we plot free-water maps using a color-scale that increases the visibility of the lower values. We can then see that the multi-shell maps are not as smoothed as the single-shell maps, although both estimations used the same Laplace-Beltrami regularization operator. As a result the multi-shell map is more detailed, allowing to better distinguish cortical structures. For this acquisition scheme, the
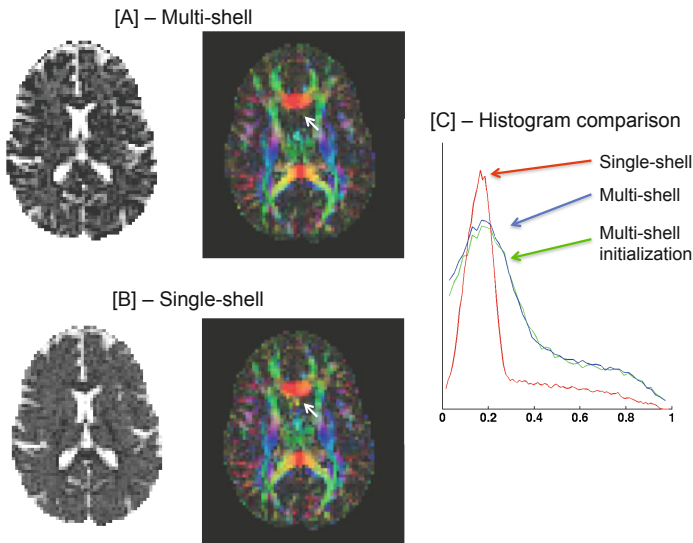

Fig. 2. Optimized multi-shell. Our multi-shell regularized fitting provides a more detailed extracellular volume and tissue tensor maps [A], comparing with the regularized single-shell fitting [B]. Small details are preserved and there are less artifacts (white arrows). The initialization f_L is very close to the final convergence as can be seen in the histogram of the images [C].

initialization, f_L is very similar to the final multi-shell free-water map. Subtle changes between the maps can be seen in the histogram comparison (Fig. 2[C]). This means that in this case f_L is a good estimator for the extracellular volume. The histogram also demonstrated that most brain voxels have extracellular volume around the 0.2 value.

Shells Added to a DTI Acquisition. This option suits cases where on-going DTI studies exist, but there is the opportunity to improve the acquisition by adding images to the existing protocol. In this case adding only lower shells is allowed, to prevent from changes in important imaging parameters such as echo time, repetition time and diffusion time that are required to accommodate the higher b-values. However, as can be seen in Fig. 3, the convergence point is no longer similar to the initialization point. The maps following regularization are smoother, but unlike the single-shell case, the details are better preserved during the regularization. The quality of the regularized images suggests that when the separability between shells is not sufficient, using the full proposed pipeline, including the regularization step is preferred.
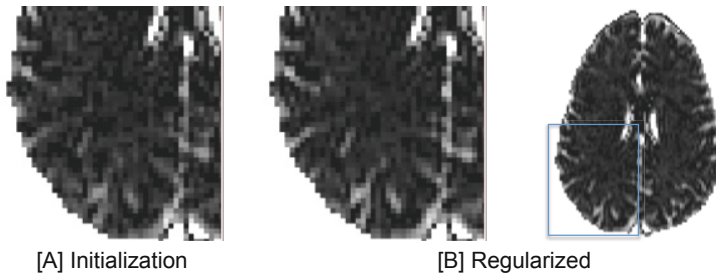

[A] Initialization   [B] Regularized

Fig. 3. Modified DTI acquisition. When there are residuals of free-water in the higher shells, the initialized free-water map [A] is different than the converged map following regularization [B]. The final map is less noisy and structures are accented. The higher quality of the regularized maps suggests that using the full proposed pipeline is preferred for this type of acquisition.

5   Discussion and Conclusions

Fitting the free-water model instead of the DTI model adds the extracellular volume as a new estimated biological parameter and provides tensor images that are more tissue specific [3,4,5,6]. We demonstrate that with a multi-shell data we can relax assumptions and dependency on regularization that is required when fitting the model from a single-shell data. We were able to estimate the extracellular volume from multi-shell data, achieving more detailed maps than single-shell maps. Although the difference is subtle, it could be important for studies of pathologies that slightly affect the extracellular space, such as neuroinflammation. For an optimized spread of a relatively small number of acquisitions in b-value shells within the DTI b-value range we can estimate the

model parameters directly from the data, avoiding regularization. Future studies could further explore optimized scheme designs, and investigate the integration with other HARDI methods that acquire multiple-shells, in which case further parameters such as the diffusivity of the isotropic compartment, can be added.

References

1. Assaf, Y., Pasternak, O.: Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. J. Mol. Neurosci. 34(1), 51–61 (2008)
2. Vos, S.B., Jones, D.K., Viergever, M.A., Leemans, A.: Partial volume effect as a hidden covariate in DTI analyses. Neuroimage 55(4), 1566–1576 (2011)
3. Jones, D.K., Cercignani, M.: Twenty-five pitfalls in the analysis of diffusion MRI data. NMR in Biomedicine 23(7), 803–820 (2010)
4. Pasternak, O., Sochen, N., Gur, Y., Intrator, N., Assaf, Y.: Free water elimination and mapping from diffusion MRI. Magn., Reson. Med. 62(3) (2009)
5. Metzler-Baddeley, C., O'Sullivan, M.J., Bells, S., Pasternak, O., Jones, D.K.: How and how not to correct for CSF-contamination in diffusion MRI. Neuroimage 59(2), 1394–1403 (2012)
6. Wang, Y., Wang, Q., Haldar, J.P., Yeh, F.C., Xie, M., Sun, P., Tu, T.W., Trinkaus, K., Klein, R.S., Cross, A.H., Song, S.K.: Quantification of increased cellularity during inflammatory demyelination. Brain 134(12), 3590–3601 (2011)
7. Metzler-Baddeley, C., Jones, D., Belaroussi, B., Aggleton, J., O'Sullivan, M.: Frontotemporal connections in episodic memory and aging: A diffusion MRI tractography study. J. Neurosci. 31(37), 13236–13245 (2011)
8. Pasternak, O., Westin, C.-F., Bouix, S., Shenton, M.E., Kubicki, M.: Free water modulation of white matter integrity measures - with application to schizophrenia. In: Proc. 19th ISMRM, p. 5309 (2011)
9. Fritzsche, K., Stieltjes, B., van Bruggen, T., Meinzer, H.P., Westin, C.F., Pasternak, O.: A combined approach for the elimination of partial volume effects in diffusion MRI. In: Proc. 20th ISMRM, p. 3548 (2012)
10. Mulkern, R.V., Haker, S.J., Maier, S.E.: On high b diffusion imaging in the human brain: ruminations and experimental insights. Magn. Reson. Imaging 27(8), 1151–1162 (2009)
11. Pierpaoli, C., Jones, D.: Removing CSF contamination in brain DT-MRIs by using a two-compartment tensor model. In: Proc. 12th ISMRM, p. 1215 (2004)
12. Pasternak, O., Sochen, N., Basser, P.J.: The effect of metric selection on the analysis of diffusion tensor MRI data. Neuroimage 49(3), 2190–2204 (2010)

Nonnegative Definite EAP and ODF Estimation via a Unified Multi-shell HARDI Reconstruction

Jian Cheng[1,2], Tianzi Jiang[1], and Rachid Deriche[2]

[1] CCM, LIAMA, Institute of Automation, Chinese Academy of Sciences, China
[2] Athena Project Team, INRIA Sophia Antipolis – Méditerranée, France
jian.cheng.1983@gmail.com

Abstract. In High Angular Resolution Diffusion Imaging (HARDI), Orientation Distribution Function (ODF) and Ensemble Average Propagator (EAP) are two important Probability Density Functions (PDFs) which reflect the water diffusion and fiber orientations. Spherical Polar Fourier Imaging (SPFI) is a recent model-free multi-shell HARDI method which estimates both EAP and ODF from the diffusion signals with multiple b values. As physical PDFs, ODFs and EAPs are nonnegative definite respectively in their domains $\mathbb{S}^2$ and $\mathbb{R}^3$. However, existing ODF/EAP estimation methods like SPFI seldom consider this natural constraint. Although some works considered the nonnegative constraint on the given discrete samples of ODF/EAP, the estimated ODF/EAP is not guaranteed to be nonnegative definite in the whole continuous domain. The Riemannian framework for ODFs and EAPs has been proposed via the square root parameterization based on pre-estimated ODFs and EAPs by other methods like SPFI. However, there is no work on how to estimate the square root of ODF/EAP called as the wavefuntion directly from diffusion signals. In this paper, based on the Riemannian framework for ODFs/EAPs and Spherical Polar Fourier (SPF) basis representation, we propose a unified model-free multi-shell HARDI method, named as Square Root Parameterized Estimation (SRPE), to simultaneously estimate both the wavefunction of EAPs and the nonnegative definite ODFs and EAPs from diffusion signals. The experiments on synthetic data and real data showed SRPE is more robust to noise and has better EAP reconstruction than SPFI, especially for EAP profiles at large radius.

1 Introduction

Diffusion MRI (dMRI) is the unique technique to explore the complex microstructure of white matter non-invasively, by modelling the diffusion of water molecules. The water diffusion is fully characterized by the diffusion Probability Density Function (PDF) called as the Ensemble Average Propagator (EAP). Under the narrow pulse assumption, the signal attenuation $E(\mathbf{q})$ is the Fourier transform of the EAP denoted by $P(\mathbf{R})$ [2]:

$$E(\mathbf{q}) = \int_{\mathbb{R}^3} P(\mathbf{R}) \exp(-2\pi i \mathbf{q}^T \mathbf{R}) d\mathbf{R} \tag{1}$$

where $\mathbf{q} = q\mathbf{u}$ is the wavevector in imaging $\mathbf{q}$-space, and $\mathbf{R} = R\mathbf{r}$ is the displacement vector in spatial $\mathbf{R}$-space, and $\mathbf{u}$ and $\mathbf{r}$ are unit vectors. Since Diffusion Tensor Imaging (DTI) cannot handle the complex fiber configuration, a category of reconstruction

methods, named as High Angular Resolution Diffusion Imaging (HARDI), were proposed to avoid the Gaussian EAP assumption in DTI [11,7,8,1,5,10]. In HARDI, EAP and two kinds of the Orientation Distribution Functions (ODFs) defined as $\Phi_0(\mathbf{r}) = \frac{1}{Z}\int_0^\infty P(R\mathbf{r})\mathrm{d}R$, $\Phi_2(\mathbf{r}) = \int_0^\infty P(R\mathbf{r})R^2\mathrm{d}R$, are normally used to infer fiber directions, where Z in $\Phi_0(\mathbf{r})$ is the normalization factor to make $\Phi_0(\mathbf{r})$ as a PDF.

Spherical Polar Fourier Imaging (SPFI) is a recent multi-shell HARDI method, which represents the signal by Spherical Polar Fourier (SPF) basis [1] and analytically obtains EAP via the Fourier dual SPF basis [5] and the ODFs via Spherical Harmonic (SH) basis [3]. Although SPFI works well for the data with low SNR and non-exponential decay [3,5], the estimated ODF/EAP may have negative values. As physical PDFs, EAPs and ODFs should be nonnegative definite in $\mathbb{R}^3$ and $\mathbb{S}^2$ respectively. However, to our knowledge the existing ODF/EAP estimation methods like the classical Q-ball Imaging (QBI) [11,7] and the recent SPFI [1,5] seldom consider this constraint. Some works considered this constraint only on the given discrete PDF samples in $\mathbb{S}^2$ for ODFs [8] and in $\mathbb{R}^3$ for EAPs [10]. However, the discrete constraint only can ensure the estimated ODF/EAP is nonnegative on the given samples, while it may be negative in other samples. Moreover it is impractical for EAPs to consider the constraint on exhaustive samples in unbounded $\mathbb{R}^3$ [10]. To our knowledge, there is still no work to estimate nonnegative definite ODFs/EAPs in the whole continuous domains.

The Riemannian framework has been proposed for tensor processing, e.g. the positive definite tensor estimation [9]. Recently the Riemannian framework has been generalized to ODF and EAP processing [4,6] by representing the square root of ODF/EAP called as the wavefunction with some orthonormal bases, and the wavefunction is calculated from the pre-estimated ODF/EAP by other methods like SPFI [1,5]. However since SPFI with the least square estimation does not consider the nonnegative constraint [1,5], the negative values of pre-estimated ODF/EAP must be forced to zero for the wavefunction estimation [6], which results in some numerical errors.

In this paper, we propose a model-free multi-shell HARDI method, named as Square Root Parameterized Estimation (SRPE), to estimate simultaneously the wavevector of EAP denoted as $\psi(\mathbf{R})$, the nonnegative EAP and ODFs from the diffusion signal samples. SRPE naturally guarantees the estimated ODF/EAP nonnegative definite in the continuous domain, not just in some discrete samples in [8,10]. The wavefunction estimated by SRPE can be used in the Riemannian framework without the numerical error introduced by negative values. Compared to SPFI, the experiments demonstrate the EAPs obtained in SRPE is more robust to noise especially at large radius.

2   Square Root Parameterized Estimation (SRPE)

2.1   Analytical Relation between the Wavefunction $\psi(\mathbf{R})$ and the Signal $E(\mathbf{q})$

SPF basis is a complete orthonormal basis which can sparsely represent Gaussian-like function with the first several basis functions [1,5], Motivated by the square root parameterization used in the Riemannian framework [4,6], we represent the wavefunction of EAP $P(\mathbf{R})$ as a linear combination of SPF basis in (2), where $G_n(R|\zeta)Y_l^m(\mathbf{r})$ is the SPF basis with the Gaussian-Laguerre function $G_n(R|\zeta)$ in radial part and the

Spherical Harmonic (SH) basis $Y_l^m(\mathbf{u})$ in spherical part [1,5], and the coefficient vector $\mathbf{c} = (c_{000}, \ldots, c_{NLL})^T$ has unit norm because $\int_{\mathbb{R}^3} P(\mathbf{R})d\mathbf{R} = 1$ [6].

$$\psi(\mathbf{R}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{nlm} G_n(R|\zeta) Y_l^m(\mathbf{r}), \quad \|\mathbf{c}\| = 1, \quad P(\mathbf{R}) = (\psi(\mathbf{R}))^2 \tag{2}$$

$$G_n(R|\zeta) = \kappa_n(\zeta) \exp\left(-\frac{R^2}{2\zeta}\right) L_n^{1/2}\left(\frac{R^2}{\zeta}\right), \quad \kappa_n(\zeta) = \left[\frac{2}{\zeta^{3/2}} \frac{n!}{\Gamma(n+3/2)}\right]^{1/2} \tag{3}$$

How to set the scale ζ will be discussed in 2.4. Please note that [6] represented $\psi(\mathbf{R})$ with Fourier dual SPF basis, while we use SPF basis. Actually different basis obtains the equivalent Riemannian framework as demonstrated in [6], however it is convenient for the analytical ODF and EAP estimation in 2.3 if $\psi(\mathbf{r})$ is represented by SPF basis.

By substituting the plane wave equation in (4) [5] into (1), where $j_\alpha(x)$ is the α-th order spherical Bessel function, the signal $E(\mathbf{q})$ in (1) can be written as a function with respect to $\mathbf{c}$ in (5), where the Fourier integration in $\mathbb{R}^3$ is separated into radial integration $I_{nn'\alpha}(q)$ and spherical integration $Q_{ll'\alpha}^{mm'\beta}$ in (7).

$$e^{-2\pi i \mathbf{q} \cdot \mathbf{R}} = 4\pi \sum_{\alpha=0}^{\infty} \sum_{\beta=-\alpha}^{\alpha} (-1)^{\alpha/2} j_\alpha(2\pi qR) Y_\alpha^\beta(\mathbf{u}) Y_\alpha^\beta(\mathbf{r}) \tag{4}$$

$$E(\mathbf{q}|\mathbf{c}) = \int_{\mathbb{R}^3} \left(\sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{nlm} G_n(R|\zeta) Y_l^m(\mathbf{r})\right)^2 e^{-2\pi i \mathbf{q}^T \mathbf{R}} d\mathbf{R}$$
$$= \sum_{nlm} \sum_{n'l'm'} \sum_{\alpha\beta} 4\pi(-1)^{\frac{\alpha}{2}} c_{nlm} c_{n'l'm'} I_{nn'\alpha}(q|\zeta) Q_{ll'\alpha}^{mm'\beta} Y_\alpha^\beta(\mathbf{u}) = \mathbf{c}^T \mathbf{K}(\mathbf{q}|\zeta)\mathbf{c} \tag{5}$$

$$\mathbf{K}_{nlm}^{n'l'm'}(\mathbf{q}|\zeta) = \sum_{\alpha=0}^{2L} \sum_{\beta=-\alpha}^{\alpha} 4\pi(-1)^{\frac{\alpha}{2}} I_{nn'\alpha}(q) Q_{ll'\alpha}^{mm'\beta} Y_\alpha^\beta(\mathbf{u}) \tag{6}$$

$$I_{nn'\alpha}(q|\zeta) = \int_0^{\infty} G_n(R|\zeta) G_{n'}(R|\zeta) j_\alpha(2\pi qR) R^2 dR, \quad Q_{ll'\alpha}^{mm'\beta} = \int_{\mathbb{S}^2} Y_l^m(\mathbf{r}) Y_{l'}^{m'}(\mathbf{r}) Y_\alpha^\beta(\mathbf{r}) d\mathbf{r} \tag{7}$$

The spherical integration $Q_{ll'\alpha}^{mm'\beta}$ is the integration of three SHs, which can be calculated by Wigner 3-j symbol. Please note that the summation over α in (5) is up to $2L$, because if $\alpha > 2L$, then $\alpha > 2L \geq l+l'$ violates the triangle inequality and $Q_{ll'\alpha}^{mm'\beta} = 0$. The radial integration in Eq. (7) can be written in (8),

$$I_{nn'\alpha}(q|\zeta) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{1.25}}{2\sqrt{q}} \int_0^{\infty} x^{1.5} \exp(-x^2) L_n^{0.5}(x^2) L_{n'}^{0.5}(x^2) J_{\alpha+0.5}(2\pi qx\sqrt{\zeta})dx \tag{8}$$

where $J_{\alpha+0.5}(x) = \sqrt{\frac{2x}{\pi}} j_\alpha(x)$ is the Bessel function. Consider $L_n^{0.5}(x) = \sum_{i=0}^{n} l_n^i x^i$, $l_n^i = (-1)^i \binom{n+0.5}{n-i}\frac{1}{i!}$, then $L_n^{0.5}(x^2) L_{n'}^{0.5}(x^2) = \sum_{i=0}^{n+n'} h_{nn'}^i x^{2i}$, $h_{nn'}^i = \sum_{j=0}^{\min(n,i)} l_n^j l_{n'}^{i-j} \delta(i-j \leq n')$ Thus the radial integration can be solved based on the property of Bessel function [5],

$$I_{nn'\alpha}(q|\zeta) = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\zeta^{0.5\alpha+1.5}\pi^{\alpha+0.5}q^\alpha}{4\Gamma(\alpha+1.5)} \sum_{i=0}^{n+n'} h_{nn'}^i \Gamma(\frac{1}{2}\alpha+i+\frac{3}{2})\,_1F_1(\frac{2i+\alpha+3}{2}; \alpha+\frac{3}{2}; -\pi^2 q^2\zeta) \tag{9}$$

where $_1F_1$ is the confluent hypergeometric function. The final quadratic relation $E(\mathbf{q}|\boldsymbol{c}) = \boldsymbol{c}^T \mathbf{K}(\mathbf{q}|\zeta)\boldsymbol{c}$ is quite compact, where the kernel $\mathbf{K}(\mathbf{q}|\zeta)$ is a $N(L+1)(L+2)/2$ dimensional symmetric matrix for each fixed $\mathbf{q}$ and ζ, and $\mathbf{K}(\mathbf{q}|\zeta)$ is independent of data $E(\mathbf{q})$.

2.2   Estimation of the Wavefunction Directly from the Signal

With the analytical relation in (5), we propose to estimate the coefficient $\boldsymbol{c}$ from the signal samples $\{E_i\}_{i=1}^{N_s}$ by minimizing the cost function $M(\boldsymbol{c})$ in (10), where the regularization matrix Λ is the diagonal matrix with the elements $\Lambda_{nlm} = \lambda_n n^2(n+1)^2 + \lambda_l l^2(l+1)^2$ which is motivated by QBI [7] and has been successfully used in SPFI [1,5,3].

$$\boldsymbol{c} = \underset{\|\boldsymbol{c}\|=1}{\arg\min} \, M(\boldsymbol{c}), \quad M(\boldsymbol{c}) = \frac{1}{2}\sum_{i=1}^{N_s}\left(\boldsymbol{c}^T\mathbf{K}(\mathbf{q}_i|\zeta)\boldsymbol{c} - E_i\right)^2 + \frac{1}{2}\boldsymbol{c}^T\Lambda\boldsymbol{c} \tag{10}$$

The minimization must be performed in a high dimensional sphere because of the constraint $\|\boldsymbol{c}\| = 1$. The Riemannian gradient $\nabla M(\boldsymbol{c})$ on the sphere is

$$\nabla M(\boldsymbol{c}) = \frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}} - \left(\boldsymbol{c}^T\frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}}\right)\boldsymbol{c}, \quad \frac{\partial M(\boldsymbol{c})}{\partial \boldsymbol{c}} = \sum_{i=1}^{N_s} 2\left(\boldsymbol{c}^T\mathbf{K}(\mathbf{q}_i|\zeta)\boldsymbol{c} - E_i\right)\mathbf{K}(\mathbf{q}_i|\zeta)\boldsymbol{c} + \Lambda\boldsymbol{c} \tag{11}$$

Then we propose a gradient descent method in (12), where $\boldsymbol{c}^{(k)}$ means $\boldsymbol{c}$ in k-th step, dt is the step size from the standard line search method, $Exp_{\boldsymbol{c}}(\boldsymbol{v})$ is the exponential map [6]. See Algorithm 1.1 for the details, where we initially set $\boldsymbol{c}^{(0)} = (1, 0, \ldots, 0)^T$ to represent a typical isotropic Gaussian EAP, and $dt_0 = 0.1$ experimentally, considering $\|\boldsymbol{c}\| = 1$. Note that $\boldsymbol{c}^{(k)}$ in each step satisfies $\|\boldsymbol{c}^{(k)}\| = 1$ thanks to the Riemannian framework [6].

$$\boldsymbol{c}^{(k+1)} = Exp_{\boldsymbol{c}^{(k)}}\left(-dt\frac{\nabla M(\boldsymbol{c}^{(k)})}{\|\nabla M(\boldsymbol{c}^{(k)})\|}\right), \quad \text{where} \quad Exp_{\boldsymbol{c}}(\boldsymbol{v}) = \boldsymbol{c}\cos\|\boldsymbol{v}\| + \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\sin\|\boldsymbol{v}\| \tag{12}$$

2.3   Estimation of the Nonnegative Definite EAP and ODFs

After the coefficient $\boldsymbol{c}$ is estimated from signal samples $\{E_i\}$, the EAP $P(\mathbf{R}) = (\psi(\mathbf{R}))^2$ is naturally nonnegative definite in $\mathbb{R}^3$, and $\int_{\mathbb{R}^3} P(\mathbf{R})d\mathbf{R} = 1$ because $\|\boldsymbol{c}\| = 1$ is forced in each estimation step. For given radius R_0, the EAP profile can be represented by product of SH basis or SH basis, considering $Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) = \sum_{\alpha\beta} Y_\alpha^\beta(\mathbf{r})Q_{ll'\alpha}^{mm'\alpha}$.

$$P(R_0\mathbf{r}) = (\psi(\mathbf{R}))^2 = \sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}\left(\sum_{nlm}\sum_{n'l'm'} c_{nlm}c_{n'l'm'}G_n(R_0)G_{n'}(R_0)Q_{nn'\alpha}^{mm'\beta}\right)Y_\alpha^\beta(\mathbf{r}) \tag{13}$$

Two kinds of ODFs $\Phi_0(\mathbf{r})$ and $\Phi_2(\mathbf{r})$ are also nonnegative definite in $\mathbb{S}^2$ because they are radial integrations of nonnegative $P(\mathbf{R})$. The ODFs can be analytically represented by product of SH basis or SH basis with the estimated coefficient $\boldsymbol{c}$. For $\Phi_2(\mathbf{r})$, we have

$$\Phi_2(\mathbf{r}) = \int_0^\infty (\psi(\mathbf{R}))^2 R^2 dR = \sum_{nlm}\sum_{n'l'm'}\left(\int_0^\infty G_n(R)G_{n'}(R)R^2 dR\right)c_{nlm}c_{n'l'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) \tag{14}$$

$$= \sum_{nlm}\sum_{l'm'} c_{nlm}c_{nl'm'}Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) = \sum_{\alpha=0}^{2L}\sum_{\beta=-\alpha}^{\alpha}\left(\sum_{nlm}\sum_{l'm'} c_{nlm}c_{nl'm'}Q_{ll'\alpha}^{mm'\beta}\right)Y_\alpha^\beta(\mathbf{r}) \tag{15}$$

Algorithm 1.1. Unified Estimation of Wavefunction, EAP and ODFs

Input: DWI samples $\{E_i\}$.

Output: Coefficient c of $\psi(\mathbf{R}|c)$ and the coefficients of EAP profile, ODFs under SH basis.

begin

    initialization: $c^{(0)} = (1, 0, ..., 0)^T, k = 0$; // typical isotropic Gaussian EAP

    repeat

        calculate $v = \nabla M(c^{(k)})$ in (11) ;

        if $\|v\| < \varepsilon_1$ **then** break;

        choose step size $dt \in (0, dt_0]$ via line search;

        $c^{(k+1)} = Exp_{c^{(k)}}(-dt\frac{v}{\|v\|})$;

        $k \leftarrow k + 1$;

    until $\frac{M(c^{(k-1)}) - M(c^{(k)})}{M(c^{(k-1)})} < \varepsilon_2$;

    $P(R_0\mathbf{r}) = $ Eq.(13), $\Phi_2(\mathbf{r}) = $ Eq.(15), $\Phi_0(\mathbf{r}) = $ Eq.(17)

end

Here we use the orthogonality $\int_0^\infty G_n(R)G_{n'}(R)R^2 dR = \delta_{nn'}$ [1]. It is clear in (15) that $\int_{\mathbb{S}^2} \Phi_2(\mathbf{r})d\mathbf{r} = \sum_{nlm} c_{nlm}^2 = 1$, because of the orthogonality of SHs. So the estimated $\Phi_2(\mathbf{r})$ from SRPE is indeed the nonnegative definite marginal EAP. For $\Phi_0(\mathbf{r})$, we have

$$\Phi_0(\mathbf{r}) = \frac{1}{Z} \int_0^\infty (\psi(\mathbf{R}))^2 \, dR = \frac{1}{Z} \sum_{nlm} \sum_{n'l'm'} \Big(\int_0^\infty G_n(R)G_{n'}(R)dR \Big) c_{nlm}c_{n'l'm'} Y_l^m(\mathbf{r})Y_{l'}^{m'}(\mathbf{r}) \quad (16)$$

$$\int_0^\infty G_n(R)G_{n'}(R)dR = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\sqrt{\zeta}}{2} \int_0^\infty \exp(-x) \sum_{i=0}^{n+n'} h_{nn'}^i x^{i-0.5}dx = \kappa_n(\zeta)\kappa_{n'}(\zeta)\frac{\sqrt{\zeta}}{2} \sum_{i=0}^{n+n'} h_{nn'}^i \Gamma(i + \frac{1}{2})$$

where $\Gamma(\cdot)$ is the Gamma function. Then $\Phi_0(\mathbf{r})$ is represented as

$$\Phi_0(\mathbf{r}) = \frac{\sqrt{\zeta}}{2Z} \sum_{\alpha=0}^{2L} \sum_{\beta=-\alpha}^{\alpha} \Big(\sum_{nlm} \sum_{n'l'm'} \kappa_n(\zeta)\kappa_{n'}(\zeta)c_{nlm}c_{n'l'm'} Q_{ll'\alpha}^{mm'\beta} \sum_{i=0}^{n+n'} h_{nn'}^i \Gamma(i + \frac{1}{2}) \Big) Y_\alpha^\beta(\mathbf{r}) \quad (17)$$

2.4 Implementation

The implementation of SRPE has two steps. The first step is to estimate coefficient vector c of $\psi(\mathbf{R})$ from signal samples $\{E_i\}$. The second step is to obtain EAP and ODFs analytically from formulae (13) (15) (17), which is independent of the first step. The whole estimation error is only in the first step, because the second step is analytical.

In SPFI, the artificial shell at $q = 0$ needs to be considered for the prior $E(0) = 1$ [5], which largely improves the results of SPFI. However, in SRPE $E(0) = 1$ is naturally satisfied because $\|c\| = 1 = \int_{\mathbb{R}^3} P(\mathbf{R})d\mathbf{R} = E(0)$ is forced in estimation, which can be seen as an advantage over SPFI, thanks to the Riemannian framework [6]. Similarly with SPFI, the scale ζ needs to be chosen in the first step. SPFI proposed to set ζ using a typical Apparent Diffusion Coefficient (ADC) value $D_0 = 0.7 \times 10^{-3} mm^2/s$ [5]. Motivated by SPFI, we propose to set ζ by two ways. The first way is to set $\zeta = 4\tau D_0$ such that the first SPF basis $G_0(R)Y_0^0(\mathbf{r}) \propto \exp(-\frac{R^2}{8\tau D_0})$ can represent typical isotropic signal $E(\mathbf{q}) = \exp(-4\pi^2\tau q^2 D_0)$ with the typical isotropic Gaussian EAP $N(\mathbf{R}|2\tau D_0)$. τ

is the diffusion time to calculate q from b value, i.e. $b = 4\pi^2\tau q^2$. In this way ζ is shared by all voxels, thus the kernels $\{\mathbf{K}(\mathbf{q}_i|\zeta)\}_{i=1}^{N_s}$ need to be calculated only once on samples $\{\mathbf{q}_i\}$. However the typical ADC value D_0 may be not appropriate for the voxels with the ADC far from D_0. Thus the second way is to adaptively estimate the isotropic tensor with ADC D_{iso} from signal $\{E_i\}_{i=1}^{N_s}$ in each voxel, which can be done by a standard least square estimation in DTI. Then we set $\zeta = 4\tau D_{iso}$. The isotropic EAP $N(\mathbf{R}|2\tau D_{iso})$ may better approximate the signal samples than the EAP $N(\mathbf{R}|2\tau D_0)$ provided by D_0. In this way ζ is adaptively set for each voxel, thus the kernel $\{\mathbf{K}(\mathbf{q}_i|\zeta)\}_{i=1}^{N_s}$ needs to be calculated for each voxel, which can be accelerated by reusing the pre-calculated $\{\mathbf{K}(\mathbf{q}_i|\zeta)\}_{i=1}^{N_s}$ with the close scale. Note that the adaptive D_{iso} can be also used in SPFI. The used special functions are implemented in GSL, which is very efficient. The computation burden is in the calculation of the kernel $\mathbf{K}$ and the summation in (11) for many times. With our C++ codes in ordinary PC, for 10000 voxels, it takes about 7 minutes with fixed scale and 20 minutes with adaptive scale.

3   Experiments

In practice we found that the ODFs estimated by many HARDI methods normally have only a small number of negative values close to zero when the SNR is very low, however even with high SNR, the negative values are serious for the estimated EAPs especially for large radius R. Thus we focus on EAP estimation in experiments.

Synthetic Data. The synthetic data were generated from mixture of tensor model [7] where two tensors cross with a given angle in $[45°, 90°]$. Three shells (b=500,1500,3000 s/mm^2) were used, 60 samples per shell. EAP profiles with radius $R_0 = 15\mu m$ were estimated by SPFI and SRPE. The Normalized Mean Square Error (NMSE) between the ground truth EAP profile $P(R_0\mathbf{r})$ and the estimated EAP profile $\widetilde{P}(R_0\mathbf{r})$ is defined as $\frac{\sqrt{\int_{\mathbb{S}^2} |\widetilde{P}(R_0\mathbf{r}) - P(R_0\mathbf{r})|^2 d\mathbf{r}}}{\sqrt{\int_{\mathbb{S}^2} |P(R_0\mathbf{r})^2 d\mathbf{r}}}$. In the noise-free experiment, the signal was generated from two tensor configurations with eigenvalues $T_1 = (1.7, 0.3, 0.3) \times 10^{-3} mm^2/s$ and $T_2 = (0.9, 0.3, 0.3) \times 10^{-3} mm^2/s$. We set $N = 2$, $L = 4$, $\lambda_l = \lambda_n = 0$ and considered both typical scale and adaptive scale for SPFI and SRPE . Fig. 1(A2,A3) recorded the NMSE when two maxima were detected. SRPE generally obtains lower NMSE and has better angular resolution than SPFI. The adaptive scale obtains lower NMSE in two methods when T_2 is used, which is because the ADC in tensor T_1 is much close to the typical D_0, while the ADC in T_2 is not. Fig. 1(A1) shows the ground truth EAP and the estimated EAPs by two methods when T_1 and crossing angle of $55°$ are used. SRPE has better angular resolution and avoids the negative values around the original point in the EAP by SPFI. Note that the EAP profile estimated by SPFI in (A1) has more than 20% negative values showed in the blue square, although only 1% points are negative and have absolute values larger than one tenth of the maximal value of the EAP profile. In the experiment with Rician noise, T_1 and adaptive scale were used. We set $\lambda_l = \lambda_n = 10^{-8}$ for SPFI suggested in [5]. Since the coefficient $\mathbf{c}$ in SRPE has different range ($\|\mathbf{c}\| = 1$) from coefficients in SPFI, in order to perform a fair comparison, we still set $\lambda_l = \lambda_n = 0$ for SRPE without any regularization. The estimation was performed for 1000 trials with
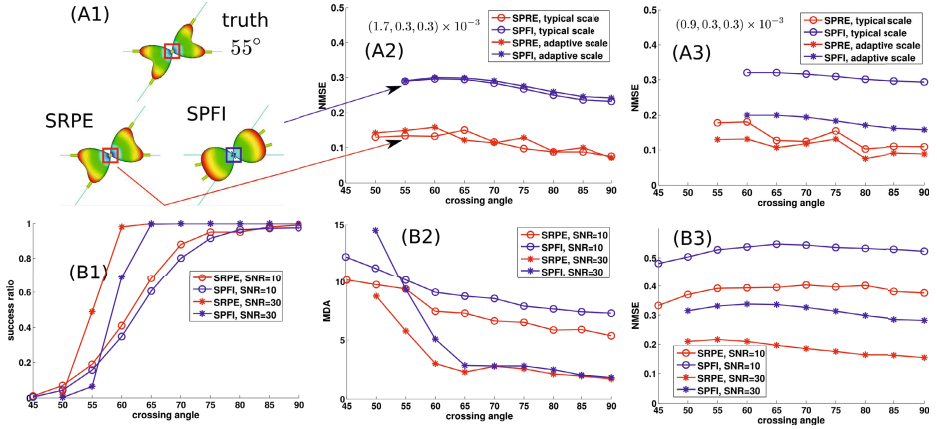
Fig. 1. A1: ground truth EAP and estimated EAPs from two methods, where the EAP by SPFI has negative values in the blue square. The long thin sticks and short thick sticks are the ground truth directions and the detected maxima respectively. A2, A3: NMSE in noise free experiment for two tensor configurations T_1, T_2. B1, B2, B3: success ratio, MDA and the mean of NMSE in the experiments with $SNR = 10, 30$.

$SNR = 10, 30$, where the success ratio was recored when two maxima were detected, the Mean Difference of Angle (MDA) was calculated in the successful trials, and the mean of NMSE was calculated over all trials. See Fig. 1(B1,B2,B3). It is clear that SRPE generally has higher success ratio, lower MDA and lower NMSE than SPFI.

Real Monkey Data. We perform SRPE and SPFI in a real monkey data with three b values ($b = 500, 1500, 3000 s/mm^2$), 30 gradients per shell. $N = 2$, $L = 4$ $\lambda_n = \lambda_l = 10^{-8}$ were set for SPFI. In order to perform a fair comparison, we set $N = 2$, $L = 4$, $\lambda_n = \lambda_l = 0$ for SRPE. See Fig. 2 for the estimated EAPs and ODFs in an enlarged area. The fifth column demonstrates that the ODFs $\Phi_2(\mathbf{r})$ by two methods are similar. It is probably because the estimated ODFs by SPFI in this area only have averagely 0.03%
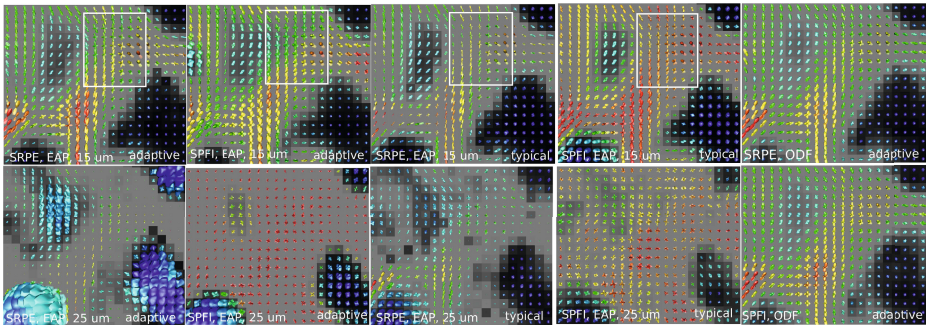
Fig. 2. The first four columns are the EAP profiles at $15, 25 \mu m$ estimated by SRPE and SPFI with adaptive and typical scales. The last column shows the ODF $\Phi_2(\mathbf{r})$ estimated by SRPE and SPFI with adaptive scale. The EAP profiles and ODFs are colored by generalized FA (GFA) [11].

negative values. The EAP profiles at radius $25\mu m$ and $15\mu m$ were estimated by two methods with both adaptive and typical scales. For both scale settings, the SPFI obtains very noisy EAP profiles at $25\mu m$, while SRPE obtains cleaner results. That's probably because the EAP profiles at $25\mu m$ by SPFI in this area have averagely more than 20% negative values, while the EAPs by SRPE are always nonnegative. Note that the EAP profiles by SRPE with adaptive scale is sharper than the EAPs obtained by typical scale, and both SPFI and SRPE obtain isotropic EAP profile in grey matter areas, which is an important advantage over other methods like QBI.

4   Conclusion

In this paper, we propose a unified model-free multi-shell HARDI method, named as SRPE, to estimate simultaneously the wavefunction of EAP, the nonnegative definite EAP and two kinds of ODFs from the diffusion signals. To the best of our knowledge, this is the first work on nonnegative definite EAP and ODFs estimation in the continuous domains, although some other methods have considered the nonnegative constraint on some given discrete samples. SRPE generalizes the positive definite tensor estimation based on the Riemannian framework for tensors to the nonnegative definite EAP and ODFs estimation by considering the Riemannian framework for EAPs. The experiments on synthetic data and real data demonstrated that the negative values happen in reconstruction methods like SPFI even without noise. This phenomenon is more serious for EAP profiles with larger radius. SRPE can improve the estimation results by avoiding the negative values, and it generally has better EAP estimation than SPFI especially for the EAP profile with large radius. The ODFs by SRPE and SPFI seem to be similar, probably because the estimated ODFs by most HARDI methods are so smooth that they seldom have negative values or have the negative values with small absolute values.

References

1. Assemlal, H.E., Tschumperlé, D., Brun, L.: Efficient and robust computation of PDF features from diffusion MR signal. Medical Image Analysis 13, 715–729 (2009)
2. Callaghan, P.T.: Principles of nuclear magnetic resonance microscopy. Oxford University Press (1991)
3. Cheng, J., Ghosh, A., Deriche, R., Jiang, T.: Model-Free, Regularized, Fast, and Robust Analytical Orientation Distribution Function Estimation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 648–656. Springer, Heidelberg (2010)
4. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: A Riemannian Framework for Orientation Distribution Function Computing. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 911–918. Springer, Heidelberg (2009)
5. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: Model-Free and Analytical EAP Reconstruction via Spherical Polar Fourier Diffusion MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 590–597. Springer, Heidelberg (2010)
6. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: Diffeomorphism Invariant Riemannian Framework for Ensemble Average Propagator Computing. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 98–106. Springer, Heidelberg (2011)

7. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, Fast and Robust Analytical Q-ball Imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
8. Goh, A., Lenglet, C., Thompson, P.M., Vidal, R.: Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 877–885. Springer, Heidelberg (2009)
9. Lenglet, C., Rousson, M., Deriche, R.: DTI segmentation by statistical surface evolution. IEEE Transactions on Medical Imaging 25, 685–700 (2006)
10. Özarslan, E., Koay, C., Shepherd, T., Blackband, S., Basser, P.: Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space mri. In: ISMRM (2009)
11. Tuch, D.S.: Q-ball imaging. Magnetic Resonance in Medicine 52, 1358–1372 (2004)

Estimation of Non-negative ODFs Using the Eigenvalue Distribution of Spherical Functions

Evan Schwab, Bijan Afsari, and René Vidal

Center for Imaging Science, Johns Hopkins University

Abstract. Current methods in high angular resolution diffusion imaging (HARDI) estimate the probability density function of water diffusion as a continuous-valued orientation distribution function (ODF) on the sphere. However, such methods could produce an ODF with negative values, because they enforce non-negativity only at finitely many directions. In this paper, we propose to enforce non-negativity on the continuous domain by enforcing the positive semi-definiteness of Toeplitz-like matrices constructed from the spherical harmonic representation of the ODF. We study the distribution of the eigenvalues of these matrices and use it to derive an iterative semi-definite program that enforces non-negativity on the continuous domain. We illustrate the performance of our method and compare it to the state-of-the-art with experiments on synthetic and real data.

Keywords: diffusion imaging, orientation distribution functions, spherical harmonics, Toeplitz matrices, eigenvalue distribution theorem.

1 Introduction

Diffusion magnetic resonance imaging (dMRI) uses the properties of water diffusion in biological tissues to reconstruct the 3-dimensional architecture of anatomical structures. Recent advances in this field, such as high angular resolution diffusion imaging (HARDI), have been able to compute the anisotropy of water molecules in the brain by measuring diffusion along multiple directions. This allows one to characterize the diffusion properties of biological tissues in terms of a probability density function on the sphere, otherwise known as the orientation distribution function (ODF).

In theory, an ODF must obey the axioms of a probability distribution, which include being non-negative and integrating to 1 over the sphere. However, existing ODF estimation methods based on a spherical harmonic (SH) representation of the ODF [1–9] do not enforce the non-negativity constraint. As a consequence, due to noise and low order SH representation, the estimated ODFs may contain negative values. This is a problem in population studies, where one is interested in applying statistical methods to ODFs to differentiate between healthy and diseased populations, which cannot be accurately done without axiomatically correct distributions. To address this problem, [10] enforces non-negativity at finitely many directions on the sphere, but ODF interpolation and registration methods may require evaluating ODFs outside discrete grids. [11] uses a non-ODF constrained

spherical deconvolution method that reduces the occurrence of negative values, but does not completely eliminate them.

In this paper, we propose an algorithm that enforces non-negativity for all directions on the sphere. To do this, we extend the relationship between continuous functions and their companion Toeplitz forms of Fourier coefficients to spherical functions. This leads to a positive semi-definiteness constraint on matrices constructed from the SH coefficients of the ODF. Since these constraints cannot be directly enforced, we study the distribution of the eigenvalues of these matrices to predict their smallest eigenvalue given an estimate of the SH coefficients. This prediction can in turn be used to improve the SH coefficients. This leads to an iterative semi-definite program (SDP) that enforces non-negativity of the ODF on the continuous domain. We illustrate the performance of our method in comparison to the state-of-the-art with experiments on synthetic and real data.

2   Estimating ODFs Using Spherical Harmonics

2.1   Standard and Modified SH Basis Representations

Recent ODF estimation methods have adopted the SH representation for HARDI signals. The (standard) SH basis elements are complex-valued functions defined as

$$Y_l^m(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{im\phi} , \quad l = 0, 1, 2, \dots, \quad -l \le m \le l, \ (1)$$

where P_l^m is the associated Legendre polynomial of degree l and order m, $\theta \in [0, \pi]$, and $\phi \in [0, 2\pi)$. In practice, the signals we want to reconstruct are real. Hence, it is more convenient to use the modified SH basis functions, which are defined as

$$Y_j = \begin{cases} \sqrt{2} Re(Y_l^{|m|}) & \text{if } -l \le m < 0, \\ Y_l^0 & \text{if } m = 0, \\ \sqrt{2}(-1)^{m+1} Im(Y_l^m) & \text{if } 0 < m \le l, \end{cases} \quad (2)$$

where $Re(\cdot)$ and $Im(\cdot)$ are the real and imaginary parts, respectively, and $j \doteq j(l,m) = \frac{l^2+l+2}{2} + m$ for $l = 0, 2, 4, \dots$ and $-l \le m \le l$. Notice that, for degree up to L, there are $R = \frac{(L+1)(L+2)}{2}$ basis elements. Often it suffices to consider the modified SH basis of degree up to $L = 4$ correlating to $R = 15$. Notice also that $\int_{\mathbb{S}^2} Y_1(\theta, \phi) = \int_{\mathbb{S}^2} \frac{1}{2\sqrt{\pi}} = 2\sqrt{\pi}$ and $\int_{\mathbb{S}^2} Y_j(\theta, \phi) = 0$ for $j > 1$. Consider a real continuous function $f : \mathbb{S}^2 \to \mathbb{R}$. Then we can write it as $f = \sum_{j=1}^{\infty} c_j Y_j$ using the modified basis in (2), where $\boldsymbol{c} = [c_j]$ are the real SH coefficients that parameterize f. This is equivalent to writing $f = \sum_{l,m} c_{l,m} Y_l^m$ using the standard SH basis in (1), where we define $\bar{\boldsymbol{c}} = [c_{l,m}]$ as the vector of standard SH coefficients. Given a real vector $\boldsymbol{c}$, we can obtain $\bar{\boldsymbol{c}}$ by the inverse mapping of $j(l,m) = \frac{l^2+l+2}{2} + m$, where $l = 0, 2, 4, \dots, \ -l \le m \le l$ with $c_{l,m} = 0$ for $l > 0$ odd. This gives the degree l_j and order m_j of Y_j and defines a one-to-one mapping between the vectors $\boldsymbol{c}$ and $\bar{\boldsymbol{c}}$. For example, $\bar{c}_{0,0} = c_1$, $\bar{c}_{2,-2} = c_2$, $\bar{c}_{2,-1} = c_3$ and so forth.

2.2   ODF Estimation Problem and Prior Work

Let S_0 be the baseline MRI signal and let $S(\theta, \phi)$ be the HARDI signal along (θ, ϕ). Following [7], we define the ODF as $p(\vartheta, \varphi) = \frac{1}{4\pi} + \frac{1}{16\pi^2} FRT\{\nabla_b^2 \ln(-\ln(\frac{S(\theta, \phi)}{S_0}))\}$, where FRT is the Funk-Radon transform, ∇_b^2 is the Laplace-Beltrami operator on $\mathbb{S}^2$, $\vartheta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$. Let $s(\theta, \phi) \doteq \ln(-\ln(\frac{S(\theta, \phi)}{S_0})) = \sum_{j=1}^{\infty} c_j Y_j(\theta, \phi)$. Since $\nabla_b^2(Y_j(\theta, \phi)) = -l_j(l_j + 1)Y_j(\theta, \phi)$, and $FRT(Y_j(\theta, \phi)) = 2\pi P_{l_j}(0)Y_j(\vartheta, \varphi)$, where $P_{l_j}(0)$ is the Legendre polynomial of degree l_j at 0, we have

$$p(\vartheta, \varphi) = \frac{1}{4\pi} + \frac{1}{16\pi^2} \sum_{j=1}^{\infty} (-2\pi P_{l_j}(0))l_j(l_j + 1)c_j Y_j(\vartheta, \varphi) = \sum_{j=1}^{\infty} \tilde{c}_j Y_j(\vartheta, \varphi), \quad (3)$$

where $\tilde{c}_1 = \frac{1}{2\sqrt{\pi}}$ and $\tilde{c}_j = -\frac{1}{8\pi} P_{l_j}(0)l_j(l_j + 1)c_j$ for $j > 1$. Therefore, to reconstruct p, it suffices to compute the SH coefficients $\tilde{\boldsymbol{c}} \doteq [\tilde{c}_j]$ that parameterize the signal s, or equivalently, the standard SH coefficients $\bar{\tilde{\boldsymbol{c}}} \doteq [\tilde{c}_{l,m}]$, such that $\int_{\mathbb{S}^2} p = 1$ and $p(\vartheta, \varphi) \geq 0$ for all $\vartheta \in [0, \pi]$ and $\varphi \in [0, 2\pi)$. To that end, assume that the HARDI signals are measured at G gradient directions $(\theta_i, \phi_i)_{i=1}^{G}$. If we use an R-dimensional approximation of $s \approx \sum_{j=1}^{R} c_j Y_j$, we have $\boldsymbol{s} \approx \mathbf{B}\boldsymbol{c}$, where $\boldsymbol{s} \doteq [\ln(-\ln(\frac{S(\theta_1, \phi_1)}{S_0})), \ldots, \ln(-\ln(\frac{S(\theta_G, \phi_G)}{S_0}))]^T$, $\mathbf{B}$ is the $G \times R$ SH basis matrix whose i-th row is $\mathbf{B}_i = [Y_1(\theta_i, \phi_i), \ldots, Y_R(\theta_i, \phi_i)]$, and $\boldsymbol{c} = [c_1, c_2, \ldots, c_R]^T \in \mathbb{R}^R$. With the above notation, we define the following ODF estimation problem:

Problem 1 (Continuous Domain ODF Estimation). Fix L and let $\boldsymbol{c} \in \mathbb{R}^R$. Solve $\min_{\boldsymbol{c}} ||\mathbf{B}\boldsymbol{c} - \boldsymbol{s}||_2^2$ s.t. $p_L(\vartheta, \varphi) \doteq \sum_{j=1}^{R} \tilde{c}_j Y_j(\vartheta, \varphi) \geq 0$ for all $\vartheta \in [0, \pi], \varphi \in [0, 2\pi)$.

Perhaps the simplest approach to recovering $\boldsymbol{c}$ is to solve the least-squares problem $\min_{\boldsymbol{c}} \frac{1}{2}||\mathbf{B}\boldsymbol{c} - \boldsymbol{s}||_2^2$, as proposed in [5]. However, disregarding the non-negativity constraints could result in negative values for $p(\varphi, \vartheta)$. To address this issue, [10] proposes to enforce the non-negativity constraints at finitely many directions $(\vartheta_i, \varphi_i)_{i=1}^{M}$, with (ϑ_i, φ_i) possibly different from (θ_i, ϕ_i).[1] To solve this problem, [10] defines the discrete ODF $\boldsymbol{p} \in \mathbb{R}^M$ whose i-th entry is $p_i = p_L(\vartheta_i, \varphi_i)$. Then $\boldsymbol{p} = \frac{1}{4\pi}\mathbf{1} + \frac{1}{16\pi^2}\mathbf{CLP}\boldsymbol{c}$, where $\mathbf{1}$ is the $M \times 1$ vector of ones, $\mathbf{C}$ is the $M \times R$ SH basis matrix whose i-th row is $\mathbf{C}_i = [Y_1(\vartheta_i, \varphi_i) \ldots Y_R(\vartheta_i, \varphi_i)]$, $\mathbf{L}$ is the $R \times R$ diagonal matrix of Laplace-Beltrami eigenvalues with $\mathbf{L}_{jj} = -l_j(l_j+1)$, and $\mathbf{P}$ is the $R \times R$ diagonal Funk-Radon transform matrix with $\mathbf{P}_{jj} = 2\pi P_{l_j}(0)$. Thus, to enforce the non-negativity of $\boldsymbol{p}$, [10] solves the optimization problem $\min_{\boldsymbol{c}} ||\mathbf{B}\boldsymbol{c} - \boldsymbol{s}||_2^2$ subject to $\mathbf{CLP}\boldsymbol{c} \geq -4\pi\mathbf{1}$. This method enforces $p_L(\vartheta_i, \varphi_i) \geq 0$ for $i = 1, \ldots, M$ but not for all $\vartheta \in [0, \pi], \varphi \in [0, 2\pi)$.

3   Estimating ODFs with Non-negativity Constraints

In this section, we propose an algorithm for solving Problem 1. In §3.1 we show that $b \doteq \min_{\vartheta, \varphi} p_L(\vartheta, \varphi) = \lim_{\ell \to \infty} \lambda_1^\ell$, where λ_1^ℓ is the smallest eigenvalue of

---

[1] Typically we consider $M = 162$ with $G \approx 100$ HARDI measurements.

a matrix T_ℓ constructed from the SH coefficients, $\bar{\bar{c}}$, of p_L. Therefore, enforcing $b \geq 0$ is equivalent to enforcing $\lambda_1^\ell \geq 0$ for all $\ell \geq L$. Unfortunately, we cannot solve an optimization problem with infinitely many constraints. Also, enforcing $\lambda_1^\ell \geq 0$ for a finite $\ell \geq L$ is a necessary but not a sufficient condition. To circumvent this problem, in §3.2 we show that the sequence $\{\lambda_1^\ell\}_{\ell \geq L}$ is decreasing and that b can be predicted by fitting a curve to the first few $\{\bar{\lambda}_1^\ell\}_{\ell=L}^N$. In §3.3, we show that given b, one can estimate $\bar{c}$ by solving an SDP. We thus propose an iterative SDP that alternates between computing $\bar{c}$ given b and predicting b given $\bar{\bar{c}}$.

3.1   Toeplitz Form Analogue for Spherical Harmonic Basis

In [12], results are developed which relate the range of the values of a function to the eigenvalues of a Toeplitz matrix (or form) constructed from finitely many Fourier coefficients of the function. One can develop analogue results for the SH coefficients. We follow [13] in constructing a matrix T_ℓ that serves as the SH analogue of the Toeplitz form in the Fourier case.

Consider a spherical function $f = \sum_{j=1}^\infty c_i Y_i$ and let $\bar{c} = [c_{l,m}]$ be its standard SH coefficients of infinite length. Let $T_\ell(f)$, $\ell = 0, 1, \ldots$, be a matrix whose rows and columns are indexed by the pair $(l_1 m_1, l_2 m_2) = (l_1(l_1+1) + m_1, l_2(l_2+1) + m_2)$, where $l_i = 0, 1, 2, \ldots, \ell$ and $-l_i \leq m_i \leq l_i$, for $i = 1, 2$. The entry of T_ℓ at position $(l_1 m_1, l_2 m_2)$ is defined as

$$T_\ell(f)_{l_1 m_1; l_2 m_2} = \sum_{l=|l_1-l_2|}^{l_1+l_2} \bar{c}_{l,m_1-m_2} G(l, l_2, l_1; m_1 - m_2, m_2, m_1), \quad (4)$$

where $G(l_1, l_2, l_3; m_1, m_2, m_3)$ is a real constant Gaunt Coefficient (See [13] Appendix A). The size of T_ℓ is $(\ell+1)^2 \times (\ell+1)^2$ because there are $(\ell+1)^2$ SH coefficients of degree less than or equal to ℓ. The following result proved in [13] relates the extremal eigenvalues of T_ℓ to the range of the values of the function f and in particular the behavior as $\ell \to \infty$.

Theorem 1. (Eigenvalue Distribution Theorem in $\mathbb{S}^2$) *Let $f(u) \in L^1(\mathbb{S}^2)$ be an absolutely integrable real valued function on the 2-sphere. Let b and B be the essential lower and upper bounds of $f(u)$, respectively and let λ_i^ℓ, $i = 1, \ldots, (\ell+1)^2$, be the ascending eigenvalues of the matrix $T_\ell(f)$. Then we have*

$$b \leq \lambda_1^\ell \leq \ldots \leq \lambda_{(\ell+1)^2}^\ell \leq B, \quad \lim_{\ell \to \infty} \lambda_1^\ell = b, \quad and \quad \lim_{\ell \to \infty} \lambda_{(\ell+1)^2}^\ell = B. \quad (5)$$

This theorem shows that the smallest eigenvalue of $T_\ell(f)$ in the limit converges to the minimum of f on $\mathbb{S}^2$. Therefore, enforcing that $\min_{\vartheta, \varphi} f(\vartheta, \varphi) \geq 0$, is equivalent to enforcing that $\lambda_1^\ell \geq 0$ for all $\ell \geq 0$.

3.2   Predicting the Smallest Eigenvalue

Since we cannot enforce infinitely many constraints, let us first understand the behavior of the smallest eigenvalue of $T_\ell(f_L)$ for $\ell \geq L$, where f_L belongs to the class of functions expressed by SH of degree up to L, i.e., $f_L = \sum_{j=1}^R c_j Y_j$.

Theorem 2. *Let $f_L = \sum_{i=1}^{R} c_i Y_i$, where $R = \frac{(L+1)(L+2)}{2}$. Let $T_\ell(f_L)$ be defined as in (4) and let λ_1^ℓ be its smallest eigenvalue. Then $\lambda_1^{\ell+1} \leq \lambda_1^\ell$ for all $\ell \geq L$.*

Proof. By [13], $T_\ell(f_L)$ is Hermitian for every $\ell \geq 0$. Then, by interlacing [14, p. 189], λ_1^ℓ will be less than or equal to the minimum eigenvalue of every principal submatrix of $T_\ell(f_L)$. Thus it suffices to show that $T_\ell(f_L)$ is a principal submatrix of $T_{\ell+1}(f_L)$ for all $\ell \geq L$. $T_\ell(f_L)$ is of size $(\ell+1)^2 \times (\ell+1)^2$ and $T_{\ell+1}(f_L)$ is of size $(\ell+2)^2 \times (\ell+2)^2$. If we restrict our matrix $T_{\ell+1}(f_L)$ to the first $(\ell+1)^2$ columns and rows, we denote this principal submatrix by $P_{\ell+1}(f_L)$. In (4) for $T_{\ell+1}(f_L)$, l_1 and l_2 each range from 0 to $\ell+1$ and so the upper limit $l_1 + l_2$ of the summation over l will range from 0 to $2\ell + 2$. But from the zero-padding construction of $T_{\ell+1}(f_L)$ we see that coefficients after 2ℓ equal 0. Thus this summation reduces to the range 0 to 2ℓ which is exactly equal to $T_\ell(f_L)$. So $T_\ell(f_L)$ is a principal submatrix of $T_{\ell+1}(f_L)$ for all $\ell \geq L$. $\square$

This result guarantees monotone decrease of λ_1^ℓ. However, to predict $\lim_{\ell\to\infty} \lambda_1^\ell$ we need the rate of convergence, which is not straightforward to analyze. In the 1D Fourier case, one can argue that this rate is roughly proportional to $\frac{1}{\ell+2}$ [12, pp. 65-67 and p. 72]. However, there are technical difficulties in carrying this argument over to $\mathbb{S}^2$. Nevertheless, in the 2D Fourier case one could argue that the rate is proportional to $\frac{1}{(\ell+2)^2}$. We use this intuition [12, 13] to obtain a good estimate of the rate of convergence. Given a sequence of eigenvalues, λ_1^ℓ, we fit functions of the form $q\frac{1}{(\ell+d)^r} + b$, with free parameters q, r, b for different values of d and small values of $\ell > L$. In addition we compare them to an exponential curve of the form $qe^{-r\ell} + b$. Table 1 shows the R^2 goodness of fit value and exponent r for different values of d averaged over 100 samples of synthetic ODFs for $\ell = 4, \ldots, 15$. We found that $d \approx 2$ and $1 \leq r \leq 2$ yield a satisfactory fit and support our intuition. In our experiments we chose $d = 2$ and let r be free.

3.3   Iterative Semi-definite Programming Optimization Algorithm

In this section, we formulate Problem 1 in terms of constraints on $T_\ell(p_L)$. Ideally, the optimization problem we want to solve is:

$$\min_{\boldsymbol{c}} \|\mathbf{B}\boldsymbol{c} - \boldsymbol{s}\|_2^2 \quad \text{s.t.} \quad T_\ell(p_L) \succeq 0, \quad \forall \ell \geq L, \tag{6}$$

where $\boldsymbol{c} \in \mathbb{R}^R$. By Theorem 1, constraining $T_\ell(p_L) \succeq 0$ for all $\ell \geq L$ enforces $p_L(\vartheta, \varphi) = \sum_{j=1}^{R} \tilde{c}_j Y_j(\vartheta, \varphi) \geq 0$ for all $\vartheta \in [0, \pi], \varphi \in [0, 2\pi)$. It is important to note here that since p_L has SH coefficients $\tilde{\boldsymbol{c}}$, $T_\ell(p_L)$ will be built from $\bar{\bar{\boldsymbol{c}}} = [\tilde{c}_{l,m}]$.

Table 1. Rate of Convergence Parameters and Performance

d	0	1	2	2.5	3	$qe^{-r\ell} + b$
R^2	0.9989	0.9990	0.9991	0.9991	0.9991	0.4522
r	0.7678	1.0162	1.2606	1.3819	1.5026	4.0142

Algorithm 1. Iterative T_ℓ Construction and SDP Optimization

Given a signal vector s, a precomputed modified SH basis matrix $\mathbf{B}$, and for a fixed L and fixed extension degree $N \geq L + 3$,

1. Initialization. Set $k = 0$.
   a. Calculate initial SH coefficient vector $c^0 = \arg\min \|\mathbf{B}c - s\|_2^2$.
   b. Construct $T_\ell(p_L^0)$ as in (4) for $\ell = L, \ldots, N$, fit curve $q\frac{1}{(\ell+2)^r} + b$ to $\{\lambda_1^{\ell,0}\}_{\ell=L}^N$, with free parameters q, r, b and set $b_0 \leftarrow b$.
2. Until b_k non-negative,
   a. Calculate γ_k, and solve (8) using SDP solver to get c^{k+1}.
   b. Construct $T_\ell(p_L^{k+1})$ as in (4) for $\ell = L, \ldots, N$, fit curve $q\frac{1}{(\ell+2)^r} + b$ to $\{\lambda_1^{\ell,k+1}\}_{\ell=L}^N$, with free parameters q, r, b, set $b_{k+1} \leftarrow b$ and set $k \leftarrow k + 1$.

Notice that we cannot solve the problem in (6), because it has infinitely many constraints. Notice also that enforcing $T_\ell(p_L) \succeq 0$ for finite ℓ does not guarantee the non-negativity of p_L. To address this issue, we propose an iterative algorithm, (see Alg. 1) that solves an approximation of (6) at each iteration. Let c^k be the estimate of c at iteration k and define $p_L^k = \sum_{j=1}^R \tilde{c}_j^k Y_j$. Let $\lambda_1^{\ell,k}$ be the minimum eigenvalue of $T_\ell(p_L^k)$ and let b^k be the prediction of $\lim_{\ell \to \infty}\{\lambda_1^{\ell,k}\}$ obtained by the curve fitting method described in §3.2. At iteration k, we solve:

$$c^{k+1} = \arg\min_c \|\mathbf{B}c - s\|_2^2 \quad \text{s.t.} \quad T_L(p_L^k) \succeq \gamma_k \mathbf{I}_{(L+1)^2}, \tag{7}$$

where $\gamma_k \doteq \lambda_1^{L,k} - b_k$ and $\mathbf{I}_{(L+1)^2}$ is the identity matrix of size $(L+1)^2$. Alg. 1 was implemented in MATLAB utilizing the Coder Toolbox to speed up construction of T_ℓ. The optimization problem was solved using CVX, a MATLAB software for specifying and solving convex programs [15, 16]. Within the CVX environment (7) is reformulated using the Shur complement of the constraint $\|\mathbf{B}c - s\|_2^2 \leq z$ as:

$$\min_c z \quad \text{s.t.} \quad \begin{bmatrix} \mathbf{I}_R & \mathbf{B}c - s \\ [\mathbf{B}c - s]^T & z \end{bmatrix} \succeq 0 \quad \text{and} \quad T_L(p_L^k) \succeq \gamma_k \mathbf{I}_{(L+1)^2} \tag{8}$$

for $z \in \mathbb{R}$. In Alg. 1, we begin with a given measured signal s and modified SH basis matrix $\mathbf{B}$ as described in §2.2. To initialize the algorithm, we fix L and start with a least squares approximation for c as in step 1a. From the initial optimal value c^0 we can calculate $\bar{\bar{c}}^0$, by the mapping in §2, in order to construct $T_L(p_L^0)$ by (4). We can optimize over any T_ℓ so we choose the smallest $\ell = L$ for computational simplicity. For a fixed integer $N \geq L + 3$ we fit the curve to $\{\lambda_1^{\ell,0}\}$ for $\ell = L, \ldots, N$ as in §3.2 in order to calculate the approximate minimum $b = b_0$ of the function p_L^0. If $b_0 < 0$ we then solve the problem in (8), which will increase the expected minimum of the sequence $\{\lambda_1^{\ell,k+1}\}$, generated from the optimal coefficients of the optimization, closer to 0. Alg. 1 repeats until we arrive at a $b_k \geq 0$, which means we have a function $p_L^k \geq 0$ on the continuous domain.

4   Experiments

We conducted a number of experiments on real and synthetic datasets to compare our continuous non-negativity (CN) method against the discrete non-negativity (DN) method [10] and the unconstrained least squares (LS) method. Our algorithm uses (DN) as initialization. We used the multi-tensor method in [4] to generate a field of 100 synthetic ODFs with 1, 2, and 3 fibers with varying angles. We added complex Gaussian noise with a signal-to-noise ratio (SNR) of 5, 10 and 20 dB. Alg. 1 takes on avg. 2-3 iterations to converge. Each iteration in `MATLAB` takes about 70s for $N = 15$ and 159s for $N = 20$, thus we choose $N = 15$ in our experiments. This is about two orders of magnitude more than the method of [10], which takes 1s per ODF. The runtime of our method could be reduced by using a more efficient implementation than `CVX`. The computational complexity of our SDP is almost $O(L^6)$. Fig. 1 offers a closeup investigation of a synthetic single fiber ODF with SNR 5 dB, using each method. Notice that CN more closely estimates the true ODF. To quantify the error, we used the Riemannian distance between ODFs, dist_{Rie} [17]. All experiments are calculated over a very fine mesh of 1.002 million points on $\mathbb{S}^2$. The left plot in Fig. 2 shows the avg. Riemannian error between the true and estimated ODFs for each SNR value. The avg. error of CN is consistently lower than that of DN. Further analysis in the righthand plot in Fig. 2 presents the avg. percentage of negative values on $\mathbb{S}^2$ for each SNR level for LS, DN, and DN with two passes of spatial regularization (DN-S) [10]. The number of negative values decreases but is not eliminated, while in CN 0% of the ODFs have negative values over the fine mesh, and in theory for all points on $\mathbb{S}^2$. Finally, we validate our methods on a 128×128 real HARDI human brain dMRI dataset measured at 94 gradient directions. Fig. 3 compares DN and CN on a sample of the real dataset. Using the same mesh as in the synthetic experi-

(a) LS   (b) DN   (c) CN   (d) True
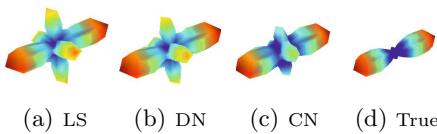
Fig. 1. Single fiber ODF with SNR 5 dB
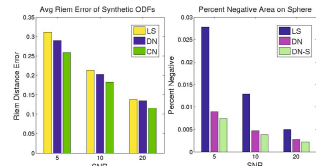
Fig. 2. Quantitative comparison
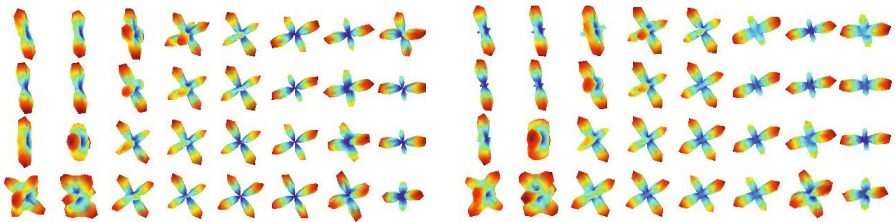
Fig. 3. Real HARDI brain ODF field. Left: DN, Right: CN

ments, we found that over 99% of the ODFs estimated by DN had on avg. 0.04% of negative values, while 0% of the ODFs estimated by CN had negative values.

5   Conclusion

We have proposed a novel method to enforce non-negativity in the estimation of ODFs. We not only eliminate the negative values of existing estimation methods, but also improve the estimated ODFs, as demonstrated with synthetic and real experiments. Future work includes using our axiomatically correct ODFs to demonstrate improvement in statistical analysis for population studies.

References

1. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine 47(6), 1083–1099 (2002)
2. Özarslan, E., Mareci, T.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magnetic Resonance in Medicine 50, 955–965 (2003)
3. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magnetic Resonance in Medicine 56(1), 104–117 (2006)
4. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical Q-ball imaging. Mag. Res. in Med. 58(3), 497–510 (2007)
5. Jian, B., Vemuri, B.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Transactions on Medical Imaging 26(11), 1464–1471 (2007)
6. Tristan-Vega, A., Westin, C.F., Aja-Fernandez, S.: Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. NeuroImage 47(2), 638–650 (2009)
7. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle. Magnetic Resonance in Medicine 64(2), 554–566 (2010)
8. Qi, L., Yu, G., Wu, E.: Higher order positive semidefinite diffusion tensor imaging. SIAM J. Imaging Sci. 3 (2010)
9. Barmpoutis, A., Vemuri, B.: A unified framework for estimating diffusion tensors of any order with symmetric positive-definite constraints. In: IEEE International Symposium on Biomedical Imaging, pp. 1385–1388 (2010)
10. Goh, A., Lenglet, C., Thompson, P., Vidal, R.: Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 877–885. Springer, Heidelberg (2009)
11. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35(4), 1459–1472 (2007)

12. Grenander, U., Szego, G.: Toeplitz Forms and their Applications. University of California Press (1958)
13. Shirdhonkar, S., Jacobs, D.: Non-negative lighting and specular object recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (2005)
14. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press (1985)
15. Grant, M., Boyd, S.: CVX: Matlab software for disciplined convex programming, version 1.21 (April, 2011), http://cvxr.com/cvx/
16. Grant, M., Boyd, S.: Graph implementations for nonsmooth convex programs. In: Blondel, V., Boyd, S., Kimura, H. (eds.) Recent Advances in Learning and Control. LNCIS, vol. 371, pp. 95–110. Springer, Heidelberg (2008)
17. Goh, A., Lenglet, C., Thompson, P., Vidal, R.: A nonparametric Riemannian framework for processing high angular resolution diffusion images and its applications to ODF-based morphometry. NeuroImage 56(1), 1181–1201 (2011)

Spatial Warping of DWI Data
Using Sparse Representation

Pew-Thian Yap and Dinggang Shen

Department of Radiology and Biomedical Research Imaging Center (BRIC)
The University of North Carolina at Chapel Hill, U.S.A
{ptyap,dgshen}@med.unc.edu

Abstract. Registration of DWI data, unlike scalar image data, is complicated by the need of reorientation algorithms for keeping the orientation architecture of each voxel aligned with the rest of the image. This paper presents an algorithm for effective and efficient warping and reconstruction of diffusion-weighted imaging (DWI) signals for the purpose of spatial transformation. The key idea is to decompose the DWI signal profile, a function defined on a unit sphere, into a series of weighted fiber basis functions (FBFs), reorient these FBFs independently based on the local affine transformation, and then recompose the reoriented FBFs to obtain the final transformed DWI signal profile. We enforce a sparsity constraint on the weights of the FBFs during the decomposition to reflect the fact that the DWI signal profile typically gains its information from a limited number of fiber populations. A non-negative constraint is further imposed so that noise-induced negative lobes in the profile can be avoided. The proposed framework also explicitly models the isotropic component of the diffusion signals to avoid undesirable reorientation artifacts in signal reconstruction. In contrast to existing methods, the current algorithm is executed directly in the DWI signal space, thus allowing any diffusion models to be fitted to the data after transformation.

1 Introduction

Spatial normalization of diffusion-weighted (DW) images often requires more than performing spatial mapping between image domains. The diffusion profile (the diffusion signals represented as a spherical function) encapsulated by each image voxel often has to be transformed to correctly align local fiber orientations. For the case of diffusion tensor imaging (DTI), this task is reduced to the reorientation of the diffusion profile based on the principal diffusion direction. For the case of high angular resolution diffusion imaging (HARDI), where the preservation of fiber crossing information is essential, the problem becomes more complicated, since the transformation has to now cater to multiple local fiber orientations in each voxel due to the existence of multiple fiber populations.

For this purpose, a decent reorientation framework was proposed by Raffelt et al. [6]. In their framework, the fiber ODF [8] is decomposed into a series of weighted spherical-harmonics-based point spread functions (PSFs), which are then reoriented individually and recomposed to form the reoriented fiber ODF. This approach was later extended in [1] for direct reorientation in the Q-space. It is further demonstrated in [1] that it is important

to take into account the isotropic component in modeling the diffusion to avoid the danger of turning an isotropic diffusion-attenuated signal profile to an anisotropic profile.

This paper proposes an algorithm for direct reorientation of the diffusion-attenuated signal profile in the Q-space by using a sparse representation framework with the DWI signal profile modeled as a combination of Watson distribution functions [10]. The proposed algorithm:

1. Avoids the computation complexity of spherical harmonics, especially that required by the associate Legendre polynomials. Although it can be argued that the spherical harmonics can be computed once and stored for subsequent computations, this strategy is generally not applicable to the case of registration, where very often the basis functions have to be computed a significant number of times, with respect to transformations that cannot be known *a priori*, as the registration algorithm iterates to refine correspondence matching;
2. Avoids the smoothing nature of spherical harmonics. When spherical harmonic basis functions of insufficient order are used, loss of sharp directional information occurs;
3. Explicitly models the isotopic diffusion component so that the isotropic content of the signal profile will not be reoriented; and
4. Incorporates an efficient non-negative L1-regularized least-squares solver, which is guaranteed to converge to the global solution in a finite number of iterative steps. This will allow us to obtain a sparse representation of the signal profile, reflecting the fact that the DWI signals at each voxel are essentially generated by a limited number of fiber populations. This is not explicitly considered in [1,6].

While employing sparse representation for diffusion modeling has been well documented (see [4] for an excellent example), the application of such framework to DWI reorientation has not been sufficiently studied. We will demonstrate that using such sparse representation framework will allow one to naturally deal with all voxels, irrespective of whether they are isotropic or highly anisotropic.

2   Approach

The proposed algorithm entails first decomposing the DWI signal profile into a series of fiber basis functions (FBFs) that are based on the Watson distribution function [10]. Given a local affine transformation, which can be computed from the local Jacobian of a deformation field estimated by any deformable registration algorithms, the FBFs are then reoriented independently and recomposed to obtain the final orientation-corrected DWI profile.

2.1   Fiber Basis Functions (FBFs)

The core of our algorithm lies in the effective decomposition of the DWI signal profile into a combination of FBFs. For better understanding of the present work, we first consider the single tensor model, with which ellipsoidal, planar and
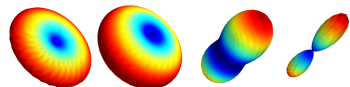
Fig. 1. The Watson distribution functions for $\kappa = -5, -1, 1, 5$

spherical directional functions can be reasonably approximated. A diffusion tensor $\mathbf{D}$ can be decomposed as $\mathbf{D} = \mathbf{UKU}^{\mathsf{T}}$, where $\mathbf{U}$ is a rotation matrix and $\mathbf{K}$ is a diagonal matrix of eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$. The eigenvalues determine the shape of the tensor. For principal diffusion along a particular direction $\boldsymbol{\mu}$ (i.e., the ellipsoidal case, $\lambda_1 \gg \lambda_2 = \lambda_3$), one can approximate the exponent in the diffusion tensor model $S(\hat{\mathbf{g}}) = S_0 \exp(-b\hat{\mathbf{g}}^{\mathsf{T}}\mathbf{D}\hat{\mathbf{g}})$ as $-b\hat{\mathbf{g}}^{\mathsf{T}}\mathbf{D}\hat{\mathbf{g}} \approx -b\lambda_1\hat{\mathbf{g}}^{\mathsf{T}}(\boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}})\hat{\mathbf{g}} = -b\lambda_1(\boldsymbol{\mu}^{\mathsf{T}}\hat{\mathbf{g}})^2$. Thus, we have the approximation $S(\hat{\mathbf{g}}) \approx S_0 \exp(-b\lambda_1(\boldsymbol{\mu}^{\mathsf{T}}\hat{\mathbf{g}})^2)$. This equation takes a form that is identical to the probability distribution function (PDF) of the bipolar Watson distribution [10]:

$$f(\hat{\mathbf{g}}|\boldsymbol{\mu}, \kappa) = C(\kappa) \exp(\kappa(\boldsymbol{\mu}^{\mathsf{T}}\hat{\mathbf{g}})^2). \tag{1}$$

The Watson distribution function hence has tensor-like properties and is especially suited for modeling the diffusion profile. The parameter $\boldsymbol{\mu}$ is a unit vector called the *mean orientation* and κ is a constant called the *concentration parameter*. $C(\kappa)$ is a normalizing constant to ensure that the density function integrates to unity over the unit sphere. Here we note that the concentration parameter κ can take both positive and negative values, giving very different shapes for the PDF. As shown in Fig. 1, negative κ values result in donut-shaped function, which is typically the shape of the diffusion profile of a fiber population with one dominant orientation. Based on this important observation, the upcoming subsections will detail how the DWI signal profile can be decomposed into a series of FBFs of different orientations for achieving the purpose of orientation correction.

2.2   Decomposing the Diffusion-Attenuated Signal Profile

Denoting the diffusion signals measured in direction $\hat{\mathbf{g}}_i$ ($i = 1, \ldots, M$) by $S(\hat{\mathbf{g}}_i)$, our aim is to represent this spherical function based on a FBF series, which in our case is realized by a combination of Watson distribution functions:

$$S(\hat{\mathbf{g}}_i) = w_0 f_0 + \sum_{j=1}^{N} w_j f(\hat{\mathbf{g}}_i|\boldsymbol{\mu}_j, \kappa) \tag{2}$$

where $\kappa < 0$ and w_j are the weights for the FBFs $f(\cdot)$. $f_0 = C(0)$ is a constant term representing the isotropic diffusion component. The directions of the FBFs, $\boldsymbol{\mu}_j$, can be set to distribute uniformly on a sphere. In matrix form, the above equation can be rewritten as $\mathbf{S} = \mathbf{Fw}$, where

$$\mathbf{S} = \begin{bmatrix} S(\hat{\mathbf{g}}_1) \\ S(\hat{\mathbf{g}}_2) \\ \vdots \\ S(\hat{\mathbf{g}}_M) \end{bmatrix}, \ \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{bmatrix}, \ \mathbf{F} = \begin{bmatrix} f_0 & f(\hat{\mathbf{g}}_1|\boldsymbol{\mu}_1, \kappa) & \ldots & f(\hat{\mathbf{g}}_1|\boldsymbol{\mu}_N, \kappa) \\ \vdots & \vdots & \ddots & \vdots \\ f_0 & f(\hat{\mathbf{g}}_M|\boldsymbol{\mu}_1, \kappa) & \ldots & f(\hat{\mathbf{g}}_M|\boldsymbol{\mu}_N, \kappa) \end{bmatrix}. \tag{3}$$

Assuming $M < N + 1$, we have a set of underdetermined linear equations, solution to which involves solving a least L2-norm problem:

$$\min_{\mathbf{w}} ||\mathbf{w}||_2 \text{ s.t. } \mathbf{Fw} = \mathbf{S}, \tag{4}$$

where $|| \cdot ||_p$ denotes the p-norm. However, to better harness the fact the DWI signals at each voxel is due to the response from a limited number of fiber populations, we compute the solution to (2) by means of a non-negative L1-regularized least-squares problem:

$$\min_{\mathbf{w}} \left\{ ||\mathbf{S} - \mathbf{Fw}||_2^2 + \beta ||\mathbf{w}||_1 \right\} \text{ s.t. } \mathbf{w} \geq \mathbf{0} \tag{5}$$

where $\beta \geq 0$ is a tuning parameter. The above problem can be solved using an active-set-based algorithm that is modified from the feature-sign algorithm presented in [5] to incorporate the non-negative constraint. The algorithm can be proven to always converge to the global optimum in a finite number of iterations.

Sparse Representation and the Isotropic Term. Determining the weight for the isotropic term by solving the least-norm problem (4) can be ambiguous. When the FBFs are distributed dense enough uniformly on a sphere, giving equal weights to all FBFs can result in an isotropic diffusion profile, hence defeating the purpose of including an isotropic term in (2) in modeling the signal profile. The sparse representation problem (5) helps avoid this pitfall by choosing the sparsest representation. In particular, in the case of an isotropic profile only w_0 will have a nonzero value.

2.3   Transformation and Recomposition

For signal profile correction in relation to spatial normalization, the directions of the FBFs, $\boldsymbol{\mu}_j$, are reoriented independently based on the local affine transformation matrix $\mathbf{A}$, i.e., $\boldsymbol{\mu}_j' = \frac{\mathbf{A}\boldsymbol{\mu}_j}{||\mathbf{A}\boldsymbol{\mu}_j||}$. Based on the reoriented FBFs, a new matrix in replacement of $\mathbf{F}$ can be computed as

$$\mathbf{F}' = \begin{bmatrix} f_0 & f(\hat{\mathbf{g}}_1|\boldsymbol{\mu}_1', \kappa) & \cdots & f(\hat{\mathbf{g}}_1|\boldsymbol{\mu}_N', \kappa) \\ \vdots & \vdots & \ddots & \vdots \\ f_0 & f(\hat{\mathbf{g}}_M|\boldsymbol{\mu}_1', \kappa) & \cdots & f(\hat{\mathbf{g}}_M|\boldsymbol{\mu}_N', \kappa) \end{bmatrix}. \tag{6}$$

The transformed DWI signal profile $\mathbf{S}'$ is finally obtained as $\mathbf{S}' = \mathbf{F}'\mathbf{w}$. Note that the isotropic component is not reoriented.

3   Experimental Results

We will first describe how the ODF and the orientations of the ODF peaks can be computed for the purpose of evaluation. We will then describe our evaluation based on simulated and *in vivo* data. For all experiments, we set $\beta = M/N$ and $\kappa = -\kappa' = -b\lambda_1$. For the *in vivo* data, λ_1 was estimated from the corpus callosum. A total of 501 orientations generated by optimizing the covering algorithm (see [2]) were used as the mean orientations of the FBFs. $T = 1281$ orientations [2] were used to locate the ODF peaks.
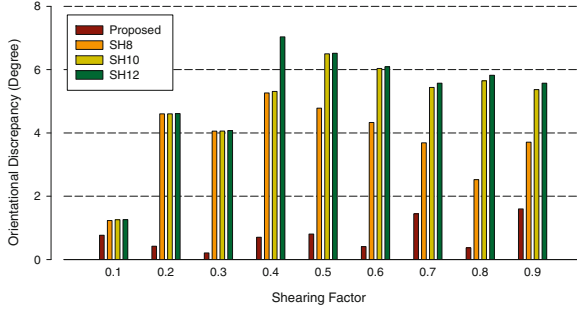
Fig. 2. Orientational discrepancy of the fiber orientations detected after reorientation with respect to the ground truth orientations
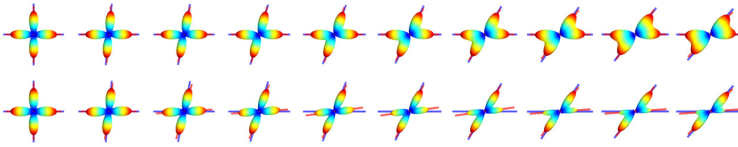
Fig. 3. Estimated fiber orientations after profile reorientation for the proposed method (top) and SH8 (bottom). Red: estimated orientations; Blue: ground truth orientations. The results from left to right correspond to shearing factor $\alpha = 0.0, 0.1, \ldots, 0.9$.

3.1   Computing the ODF and the ODF Peaks

The ODF associated with a Watson distribution function can be written as [7] $\mathcal{O}[f(\hat{\mathbf{g}}_i|\boldsymbol{\mu}_j, \kappa)] \propto f(\hat{\mathbf{g}}_i|\boldsymbol{\mu}_j, \kappa')$ with $\kappa' > 0$. We can hence write $\mathcal{O}[S(\hat{\mathbf{g}}_i)] = w_0 f_0 + \sum_j^N w_j f(\hat{\mathbf{g}}_i|\boldsymbol{\mu}_j, \kappa')$. A concentration $\kappa' = -\kappa/2$ will give results similar to that derived in [7], which is based on the method suggested by Tuch [9]. A larger value of κ' will give results closer to the sharper fiber ODF [8]. For simplicity, we used $\kappa' = -\kappa$ for all experiments. To extract the orientations of the ODF peaks, which represent the local fiber orientations, the following steps were performed:

1. Sample the ODF with sufficient density at orientations $\hat{\mathbf{w}}_1, \ldots \hat{\mathbf{w}}_T$.
2. Remove orientations associated with ODF values less than the mean value.
3. Locate orientations with values greater than their neighboring orientations.
4. Compute the mean orientations of the orientations in the neighborhood of the orientations with the maximal values. This can be done by computing the eigenvector corresponding to the largest eigenvalue of the dyadic tensor $\mathbf{D}_{\text{dyadic}}(\hat{\mathbf{w}}_i) = \frac{1}{|\mathcal{N}(\hat{\mathbf{w}}_i)|} \sum_{\mathbf{v} \in \mathcal{N}(\hat{\mathbf{w}}_i)} \mathbf{v}\mathbf{v}^{\mathrm{T}}$ for $\hat{\mathbf{w}}_i$ satisfying $\mathcal{O}[S(\hat{\mathbf{w}}_i)] > \mathcal{O}[S(\mathbf{v})], \forall \mathbf{v} \in \mathcal{N}(\hat{\mathbf{w}}_i)$. $\mathcal{N}(\hat{\mathbf{w}}_i)$ denotes the neighborhood of $\hat{\mathbf{w}}_i$. Return the mean orientations as the output.
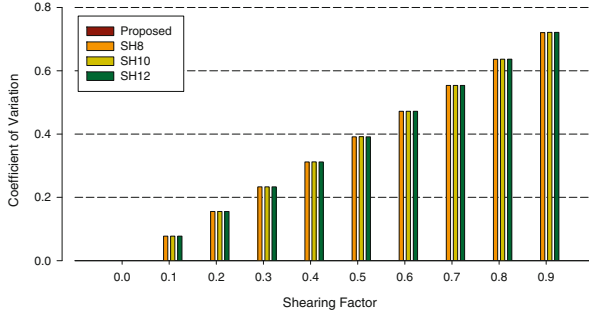
Fig. 4. Preservation of isotropy after reorientation. A greater c.v. value indicates higher anisotropy. Note that the values given by the proposed method is consistently zero for all cases; therefore, they are not visible.
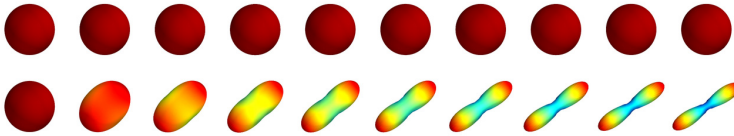
Fig. 5. Reorientation results for the proposed method (top) and SH8 (bottom) using shearing factor $\alpha = 0.0, 0.1, \ldots, 0.9$. With the proposed method, the isotropy of the profile is faithfully preserved.

3.2   Simulated Data

Assuming 2 crossing fiber populations, we used a mixture of diffusion tensors to generate a diffusion profile representing a fiber crossing for the evaluation of the proposed method. Each fiber population is represented by a tensor with $\lambda_1 = 5 \times 10^{-3} \text{mm}^2/\text{s}$, $\lambda_2 = \lambda_3 = 5 \times 10^{-4} \text{mm}^2/\text{s}$ and $b = 1000 \text{s}/\text{mm}^2$. The (120) gradient directions were taken from the *in vivo* dataset. One tensor is oriented in the horizontal (x-axis) direction and the other in the vertical (y-axis) direction.

Reorientation Accuracy. The diffusion profile was sheared in the horizontal direction using the transformation matrix $\mathbf{A} = [1 \; \alpha \; 0; \; 0 \; 1 \; 0; \; 0 \; 0 \; 1]$, where α is the *shearing factor*, increment of which will result in a greater degree of shearing. We set $\alpha = 0.1, 0.2, \ldots, 0.9$. The ground-truth orientations were computed by reorienting directly the orientations of the individual tensors, i.e., $[1, 0, 0]$ and $[0, 1, 0]$. We evaluated the accuracy of the reorientated diffusion profile by comparing fiber orientations detected from it with respect to the ground truth. Assuming that $\mathbf{U}$ is the set of ground truth orientations and $\mathbf{V}$ is the corresponding set of estimated orientations, the *orientational discrepency* is defined as $\frac{1}{2} \left[\frac{1}{|\mathbf{U}|} \sum_{\mathbf{u} \in \mathbf{U}} \min_{\mathbf{v} \in \mathbf{V}} d_\theta(\mathbf{u}, \mathbf{v}) + \frac{1}{|\mathbf{V}|} \sum_{\mathbf{v} \in \mathbf{V}} \min_{\mathbf{u} \in \mathbf{U}} d_\theta(\mathbf{v}, \mathbf{u}) \right]$, where $d_\theta(\mathbf{u}, \mathbf{v})$ gives the angle difference between $\mathbf{u}$ and $\mathbf{v}$, i.e., $d_\theta(\mathbf{u}, \mathbf{v}) = \cos^{-1}(|\mathbf{u} \cdot \mathbf{v}|)$. The absolute value is taken since diffusion is assumed to be antipodal symmetric.

The results are shown quantitatively in Fig. 2 and qualitatively in Fig. 3. The results generated using Raffelt et al.'s method [6], applied directly to the ODF using spherical harmonics up to order 8 (SH8), 10 (SH10), and 12 (SH12), are also included for comparison. Note that for the proposed algorithm the ODFs were computed based on the reoriented DWI signal profiles, whereas for the comparison method reorientation was performed directly on the ODF. The results indicate that the proposed method 1) yields lesser error in the estimated orientations, and 2) affects less the diffusion profile in the horizontal direction, which is to be expected, since shearing is applied in the horizontal direction.
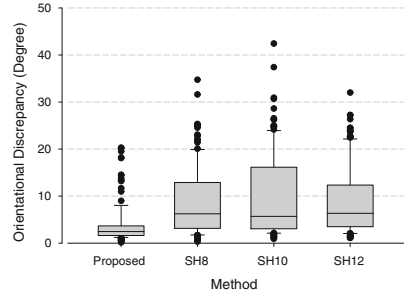
Fig. 6. Reorientation accuracy evaluated using *in vivo* data

Preservation of Isotropy. We also evaluated the proposed method on whether it can successfully preserve the isotropy of an isotropic profile. For this purpose, we measured the anisotropy of the reoriented ODFs using the coefficient of variation c.v. $= \frac{\text{std}(\mathcal{O}[\mathbf{S}])}{\langle \mathcal{O}[\mathbf{S}] \rangle}$, recalling that $\mathcal{O}[\cdot]$ is the ODF operator previously defined. A larger c.v. value indicates a greater degree of anisotropy.

We generated an isotropic diffusion profile with constant signal magnitude $\exp(-b\lambda)$, where $b = 1000\text{s/mm}^2$ and $\lambda = 5 \times 10^{-3}\text{mm}^2/\text{s}$, in all directions. This profile was then subject to the different degree of shearing identical to the experiment using simulated data. The c.v. values of the reoriented ODFs were then measured as an indication of whether the reorientation algorithm unnecessarily distorts the originally isotropic profile. The results, shown in Fig. 4, demonstrate the importance of explicitly modeling the isotropic diffusion component. Neglecting this will cause the originally isotropic profile to become anisotropic after reorientation, which cannot be physically true. Visual results for comparison are shown in Fig. 5, where it can be seen that the distortion for the representative case of SH8 is quite apparent.

3.3   Real Data

Diffusion-weighted images were acquired for an adult subject using a Siemens 3T TIM Trio MR Scanner with an EPI sequence. Diffusion gradients were applied in 120 non-collinear directions with diffusion weighting $b = 2000\text{s/mm}^2$. The imaging matrix was 128×128 with a rectangular FOV of $256 \times 256\text{mm}^2$. 80 contiguous slices with a slice thickness of 2mm covered the whole brain.

We extracted DWI signal profiles from the voxels located in the pons, since this location of the brain was found to contain a significant amount of fiber crossings [3]. The profiles were randomly transformed using the matrix $\mathbf{A} = [1 \ \alpha \ 0; \ 0 \ 1 \ 0; \ 0 \ 0 \ 1]\,\mathbf{R}$, where $\mathbf{R} = \mathbf{R}_x\mathbf{R}_y\mathbf{R}_z$ is composed by matrices of rotations around the x, y, and z axes. The local orientations prior to profile reorientation were estimated and transformed using the corresponding matrix $\mathbf{A}$ to serve as ground truth for evaluation of the

reorientation algorithms. The reorientation accuracy was measured by computing the orientational discrepancy of the estimated orientations with respect to the ground truth. The results, shown in Fig. 6, again confirms that the proposed method yields markedly improved results.

4   Conclusion

We have presented in this paper a novel algorithm for the transformation of raw DWI data directly in the Q-space. The algorithm takes into account of the isotropic diffusion component and can therefore be applied to any voxels without requiring explicitly masking out gray matter and cerebospinal fluid voxels. The capability of working directly with the diffusion signal profiles implies that the transformed outcome will allow the plethora of diffusion models to be fitted after the fact. It is not difficult to envision that future works involving registration, segmentation, and voxel-based analysis using diffusion-weighted images will benefit fundamentally from the current work.

References

1. Dhollander, T., Van Hecke, W., Maes, F., Sunaert, S., Suetens, P.: Spatial transformations of high angular resolution diffusion imaging data in Q-space. In: Workshop on Computational Diffusion MRI (CDMRI), Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 73–83 (2010)
2. Hardin, R.H., Sloane, N.J.A., Smith, W.D.: Tables of spherical codes with icosahedral symmetry, http://www2.research.att.com/~njas/icosahedral.codes/
3. Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J.: Estimating the number of fiber orientations in diffusion MRI voxels: a constrained spherical deconvolution study. In: ISMRM 2010 (2010)
4. Jian, B., Vemuri, B.C.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE Transactions on Medical Imaging 26(11), 1464–1471 (2007)
5. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS, pp. 801–808 (2007)
6. Raffelt, D., Tournier, J.D., Fripp, J., Crozier, S., Connelly, A., Salvado, O.: Non-linear spatial normalisation of high angular resolution diffusion imaging data using fiber orientation distributions. In: Workshop on Diffusion Modelling and the Fibre Cup (DMFC), Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 73–83 (2009)
7. Rathi, Y., Michailovich, O., Shenton, M.E., Bouix, S.: Directional functions for orientation distribution estimation. Medical Image Analysis 13(3), 432–444 (2009)
8. Tournier, J.D., Calamante, F., Gadian, D.G., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23(3), 1176–1185 (2004)
9. Tuch, D.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)
10. Watson, G.: Equatorial distributions on a sphere. Biometrika 52, 193–201 (1965)

Tractography via the Ensemble Average Propagator in Diffusion MRI

Sylvain Merlet[1], Anne-Charlotte Philippe[1],
Rachid Deriche[1], and Maxime Descoteaux[2]

[1] Athena Project-Team, INRIA Sophia Antipolis - Méditerranée, France
[2] Sherbrooke Connectivity Imaging Laboratory (SCIL), Canada

Abstract. It's well known that in diffusion MRI (dMRI), fibre crossing is an important problem for most existing diffusion tensor imaging (DTI) based tractography algorithms. To overcome these limitations, High Angular Resolution Diffusion Imaging (HARDI) based tractography has been proposed with a particular emphasis on the the Orientation Distribution Function (ODF). In this paper, we advocate the use of the Ensemble Average Propagator (EAP) instead of the ODF for tractography in dMRI and propose an original and efficient EAP-based tractography algorithm that outperforms the classical ODF-based tractography, in particular, in the regions that contain complex fibre crossing configurations. Various experimental results including synthetic, phantom and real data illustrate the potential of the approach and clearly show that our method is especially efficient to handle regions where fiber bundles are crossing, and still well handle other fiber bundle configurations such as U-shape and kissing fibers.

1 Introduction

At the current resolution of diffusion-weighted (DW) magnetic resonance imaging (MRI), research groups agree that there are between one and two thirds of imaging voxels in the human brain white matter that contain fibre crossing bundles [8]. We know that in these locations, the diffusion is non-Gaussian and the diffusion tensor (DT) [3] is limited due to its intrinsic Gaussian diffusion assumption. Hence, DT-based tractography algorithms can follow false tracts and produce unreliable tracking results. To overcome limitations of the DT, new HARDI techniques have been proposed to estimate the diffusion orientation distribution function (ODF) [10,5,1,9] of water molecules. These HARDI techniques were developed to deal with non-Gaussian diffusion processes and because the maxima of the ODF are aligned with the underlying fibre populations, deterministic and probabilistic ODF-based tractography algorithms have been proposed [5] that outperform classical DT-based tractography algorithms. In this paper, we mainly advocate the use of the Ensemble Average Propagator (EAP) instead of the ODF for tractography in dMRI and propose an original and efficient streamline EAP-based tractography algorithm that outperforms the classical ODF-based tractography, in particular in the regions that contain complex fiber crossing congurations.

2   Analytical Signal Estimation and Diffusion Features

Recent works [2,7,4] propose to analytically estimate the normalized diffusion signal $E(q\mathbf{u})$ as a linear combination of 3D functions Ψ_{nlm},
$E(q\mathbf{u}) = \sum_{n=0}^{N} \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{nlm} \Psi_{nlm}(q\mathbf{u})$, where q is the norm of the effective gradient, $\mathbf{u}$ a unitary vector, and n,l are respectively the radial and angular order of the associated 3D function.

In [2,7,4] the authors also propose analytical formulas to estimate some diffusion features. In particular, they enable to compute the EAP P at any radius R, i.e. $P(R\mathbf{r})$ with $\mathbf{r}$ a 3D unit vector. In our paper, we are especially interested in this feature. Moreover, they give analytical formulas to estimate the normalized ODF Υ expressed as the integration of the EAP over a solid angle [1,9], $\Upsilon(\mathbf{r}) = \int_0^\infty P(R.\mathbf{r})R^2 dR$. The ODF will be used, as comparison, in order to perform the well-known ODF based streamline algorithm [5]

3   Motivations and Challenges

Here, we motivate the use of the EAP instead of the ODF in order to characterize fiber orientations. Several challenges arise when dealing with the whole EAP information that must be taken into consideration. We illustrate this section by an example confronting ODF and EAP based estimation of the fiber orientations.

The ODF is known to well represent the angular structure of the diffusion phenomenon. However, it has some limitations. First, the radial integration over a solid angle is done all over the EAP, and thus is likely to catch unwanted artifact that lead to false orientation detection. The next problem arises when fiber bundles with large differences in anisotropy are crossing. We illustrate this with a synthetic data example. We consider two crossing fiber bundles, a high anisotropy fiber bundle (the curved one) and a low anisotropy fiber bundle (the diagonal one). We estimate the ODF $\Upsilon(\mathbf{r})$ and the EAP $P(R\mathbf{r})$ with radii $R = 5, 10, 15, 18\mu m$ via the SHORE method [7]. Then, we extract the maxima of each spherical function using a discrete approach. These maxima are represented by the red line. Because the diffusion in direction of the curved bundle highly predominates, we don't manage to resolve this crossing region using the ODF Υ (see Fig. 1.a). However, we see that the EAPs manage to catch some angular information where the ODFs fail. In particular, the whole crossing structure is caught at a radius $R = 18\mu m$ (see Fig. 1.e).

The EAP has also some disadvantages. Even, if it can catch more complex structure than the ODF, we need to know where the significant information is localized. Considering the EAP is near isotropic for low radii, and becomes more anisotropic for higher radii, it would be natural to consider only high EAP radii. However, this is not done for several reasons : 1) a physical reason and 2) a technical reason. First,the trajectory of water molecules can be described by a random walk, i.e. each molecule follows a random path in a 3D space. However, when looking at a set of molecules, one can see that, in a free medium, the averaged displacement follows a Gaussian distribution. It means that most of the
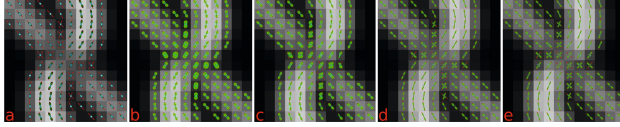
Fig. 1. ODF $\Upsilon(\mathbf{r})$ (a) and the EAP $P(R\mathbf{r})$ with radii $R = 5, 10, 15, 18\mu m$ (respectively b, c, d, e) estimated via the SHORE method in a synthetic crossing configuration, and extracted maximas. The curved fiber bundle has a high anisotropy and the diagonal bundle a low anisotropy.

molecules remain confined around their initial positions. At the scale of a voxel, the EAP represents this averaged displacement of molecules and because of its Gaussian particular nature, the EAP values decrease while moving away from its center until falling to zero and, then, the EAP return to an isotropic state. Beside physical reasons, the maximum b-value authorized by the acquisition protocol, makes the signal band limited i.e. the high frequency component are discarded. It results in a smoothing of the EAP and a loss of details. This phenomenon emphasizes the return to an isotropic state.

Because of these two reasons, we know the significant angular information is localized away from the EAP center but not too far because of the signal decrease. Beside the difficulty to find an optimal radius, we know as well that the water molecule diffusion is likely to vary between different voxel, and the angular information we want to catch is not always localized at the same radius. Therefore, we cannot consider the EAP at only one radius for the whole tractography. Hence, the challenge is to find the optimal radius R at each voxel, on which the EAP $P(R\mathbf{r})$ optimally represents the orientation of the current fiber.

4   EAP Based Tractography

The main problem in recovering the fiber bundle orientations, comes from the fact that we don't know at which EAP radius R is localized the information. The challenge, as explained in section 3 is to find the optimal EAP radius R, on which we are able to catch the orientation of the current fiber.

We could proceed as follow : Starting from an initial radius, we authorize to switch to the radius R where the local maxima of $P(R\mathbf{r})$ has the largest similarity with the current fiber direction. However, if we force the fiber to find a maxima direction similar to the current fiber direction, we would take the risk to always follow the same direction (not suitable for curved fibers). This is especially true if false local maxima are detected due to noise. Our idea is to encourage the switching between close radii by inserting a penalty function, which penalizes the radii far from the current radius. It enables to obtain consistent fibers, which are more robust to noise. Then, we define the penalty function as $w_{R_c}(R)$, which measure the amount of penalty between the current radius R_c and the radius R. We propose to use a exponential penalty function, i.e. $w_{R_c}(R) = \exp(-\beta \times \frac{|R_c - R|}{R_c})$ where β is a scale parameter.

Before continuing with the complete tracking algorithm, we need to clarify the fiber orientation detection step. Considering that we have an estimation of the EAP P, we define a procedure "ExtractAngularInformation", whose the input is $\mathbf{v}_c \in \mathbb{R}^3$ (the current fiber direction), $R_c \in \mathbb{R}$ (the current EAP radius), and $\beta \in \mathbb{R}$ (the penalty scale parameter), and returns the next direction $\mathbf{v}_n \in \mathbb{R}^3$ and the corresponding EAP radius $R_n \in \mathbb{R}$. We also need a function Extractmax(P), which return the list l_R of directions $l_R(j)$ along each maxima of $P(R\mathbf{r})$. The procedure is described in Alg. 1.

Algorithm 1. ExtractAngularInformation

Require: $\mathbf{v}_c \in \mathbb{R}^3$, $R_c \in \mathbb{R}$, $\beta \in \mathbb{R}$, the EAP P
  1. Compute $w_{R_c}(R) = \exp(-\beta \times \frac{|R_c - R|}{R_c})$ for $R \in \mathbb{R}$
  2. Compute $l_R =$ extractMax(P)
  3. Set $R^*, j^* = argmax_{R,j} \langle l_R(j) \cdot v_c \rangle \times w_{R_c}(R)$
  4. $R_n = R^*, v_n = l_{R^*}(j^*)$
  return R_n, v_n

Now we have defined a way to detect fiber orientations, we describe our deterministic tractography algorithm based on the EAP. For the purpose we extend the streamline method based on multiple ODF maxima [5]. At first, we denote p(s) as the curve parameterized by its arc-length. This curve can be computed as a 3D path adapting its tangent orientation locally according to vector field v. Hence, for a given starting point p_0, we solve $p(t) = p_0 + \int_0^t \mathbf{v}(p(s))ds$. The integration is typically performed numerically with Euler or Runge-Kutta schemes of order 2 or 4. In the Euler case, we have the discrete evolution equation $p_{n+1} = p_n + \mathbf{v}(p_n)\Delta_s$, where Δ_s is a small enough step size to obtain subvoxel precision. A continuous linear, cubic, spline interpolation of the vector field can be done at each step for the subvoxel points. For our algorithm, we need a starting seed $p_0 \in \mathbb{R}^3$, a starting radius R_0, a curvature threshold $t_\theta \in [0, 90°]$. The algorithm is described in Alg. 2. For the rest of the paper, ODF-tract refers to the ODF based streamline tractography [5], and EAP-tract refers to our EAP based tractography algorithm (Alg. 2).

Algorithm 2. EAP based fibers tracking algorithm

Require: $p_0 \in \mathbb{R}^3$, $R_0 \in \mathbb{R}$, $t_{fa} \in [0, 1]$, $t_\theta \in [0, 90°]$
  1. Estimate field of analytical EAP estimation P
  2. $\mathbf{v}_{p_0} = argmax_\mathbf{u} P_{p_0}(R_0\mathbf{u})$
  3. Set $i = 0, R_i = R_0$ and **do**
    $p_{i+1} = p_i + \Delta \mathbf{v}_{p_i}$
    Update the fiber f according to p_i
    $R_{i+1}, \mathbf{v}_{p_{i+1}} = ExtractAngularInformation(\mathbf{v}_{p_i}, P_{p_i}, R_i, \beta)$
    if $\frac{\langle \mathbf{v}_{p_i} \cdot \mathbf{v}_{p_{i+1}} \rangle}{\|\mathbf{v}_{p_i}\| \|\mathbf{v}_{p_{i+1}}\|}$ **then** stop;
    $i = i + 1$
  return f

5   Experiments

Synthetic data : We start by validating our method on synthetic data. For this purpose, we consider the EAP and ODF of Fig. 1. In this example, two bundles, a high anisotropy fiber bundle (the curved bundle) and a low anisotropy fiber bundle (the diagonal bundle) are crossing. Let us remind that, the EAP and ODF were estimated via the SHORE method. We first compare the EAP-tract and ODF-tract methods in case of noiseless measurements (see fig. 2 a,b). In both algorithms, we set the step $\Delta_s = 0.5$ and the curvature threshold $t_\theta = 75°$. The scale parameter of the penalty function in Alg. 1 is set to $\beta = 0.5$. In Fig. 2.a, we see that the diagonal fibers from the ODF-tract method fail to pass through the crossing region because the diffusion constrained by the curved fiber predominates in this area. This is not surprising while looking at Fig. 1.a, in which the ODF maxima are not sufficiently large to detect the diagonal fibers in the crossing area. However, if we look at Fig. 1.b,c,d,e, the maxima extracted from the EAP at the different radii, catch the crossing structure. Then, exploiting the angular information at any radius, our EAP-tract method efficiently resolves the crossing region where the ODF-tract fails. Only, few fibers do not follow the right path, whereas all the fibers from ODF-tract follow the wrong path. We see in the next experimental section that our method enables, as well, to recover fibers from other configurations such as U-shape and kissing. We also present EAP-tract results measurements contaminated with Rician noise with SNR=30,20,10 to the measurements (respectively Fig. 2.c,d,e). Nearly all the fibers pass through the crossing area, which shows the robustness of the EAP-tract to noise. On this example, we also want to show the robustness to the choice of the initial radius R_0 used to start the EAP tract algorithm. Hence, we launch the EAP-tract method with five different initial radii, i.e. $R_0 = 5, 10, 15, 20, 25\mu m$. Then we plot in Fig. 4-left the evolution of radii each track follows during the tractography(R_i in Alg. 2). In Fig. 4-left we perceive the crossing region, when all the curves converge, in the notified area (red circle). The radii from all the track converge approximately to the range $[16 - 18\mu m]$. This is not surprising, if we look at figure 1.b,c,d,e, where the whole angular structure appears at radius $18\mu m$. Via this example, we see that our EAP-tract method is robust to the choice of the initial radius.
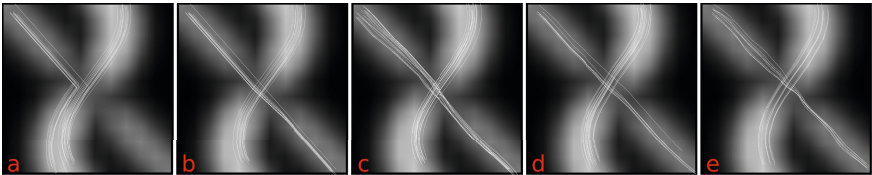
Fig. 2. Tracking on a synthetic crossing region. a,b) are respectively the ODF-tract and EAP-tract in case of noiseless measurements. c,d,e) EAP-tract with measurements contaminated with rician noise (respectively with a SNR=30,20,10).
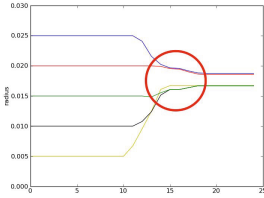
	ODF		EAP	
	F.P.	T.P.	F.P.	T.P.
seed 1	4	0	3	17
seed 8	0	21	0	12
seed 7	17	2	2	13
seed 11	5	1	0	4
seed 14	0	22	0	25
seed 15	11	5	5	12

Fig. 3. Left : Convergence of the radius in the EAP-tract method on synthetic data tractography. **Right** : Quantitative results associated to the phantom tracking.

Phantom data : We also validate our method on the phantom data used in the contest, the "fiber cup" [6] in MICCAI 2009. The ground truth is available and enables to correctly compare tractography algorithms. This phantom contains realistic fibers configurations such as crossing, kissing and bending.

Data pre-processing was allowed for the contest. However, in our experiments, we consider the raw data set. We perform our validation as following : 1) We choose a starting voxel v within the seed proposed for the contest. 2) We launch our algorithm on N_s points taken at random in the voxel v. This step results in N_s fiber tracks. 3) Within these N_s fibers, we keep only the fibers with a minimum length ℓ_{min}, and discard the fibers that do not stop outside the phantom. 4) Within the remaining fibers, we count : a) the number of fibers following the ground truth, i.e. the true positive (TP); and b) the number of fibers following another track, i.e. the false positive (FP). We launch the validation with $N_s = 30$, $l_{min} = 20$. We use the same algorithm parameters than in sec. 5. The EAP field is estimated via SHORE method. Fig. 4 shows the fiber tracks for 6 configurations corresponding to several seeds given in the fiber cup: A kissing (Purple fibers, seed 8), a U-shape bundle with high curvature (Blue light fibers, seed 14), and four different crossing configurations (Blue curve, seed 1; green curve, seed 7; yellow curve, seed 11; red curve, seed 15). We also write in Tab. 4-right, the number of true and false positive (respectively TP and FP) for each configuration of the two compared methods (EAP-tract and ODF-tract).
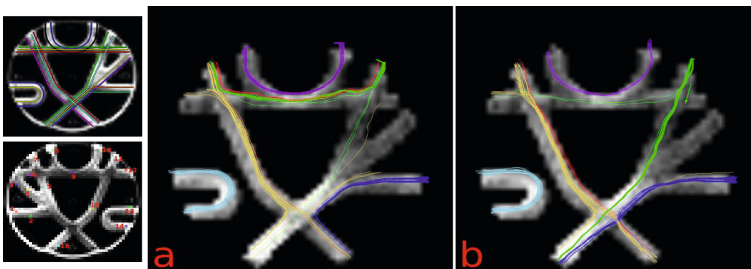
Fig. 4. Phantom of the "Fiber cup". a) ODF-tract fibers, b) EAP-tract fibers. On the left we show the ground truth fiber bundles and the spatial position of the seeds.

First, we note that the EAP-tract fibers are much in-line with the ground truth than the ODF-tract fibers. The ODF-tract method has difficulties to go through the crossing region whereas the EAP-tract method well resolve these crossing regions. These qualitative remarks are confirmed by quantitative results, while considering the Tab. 4-right. All the fibers involved in crossing bundles (seed 1,7,11,15) have a large number of FP compared to the number of TP using the ODF tract method, whereas most of the fibers from the EAP-tract method are TP. In particular, 90% of the ODF-tract fibers from seed 7 do not follow the right path, whereas only 10% of the EAP-tract fibers from seed 7 do not follow the right path. This improvement of the EAP tract over the ODF tract is huge considering the crossing region. In our study we also consider an U-shape (seed 11, light blue curve) and a kissing configuration (seed 8, purple curve). There is no large difference between the two methods for the U-shape configuration. Both efficiently handle high curvature fibers. Considering the kissing case (seed 8, purple curve), there is a small advantage of the ODF-tract method. However, we do not see any FP in both method, which means they can handle kissing fiber configurations efficiently.

To conclude this section, we see that the EAP-tract method efficiently resolve crossing fibers where the ODF-tract fails. Considering other fiber bundle configurations (U-shape and kissing), both methods are equivalent. Hence the advantage of the EAP-tract method lies in its great efficiency to pass trough crossing regions.
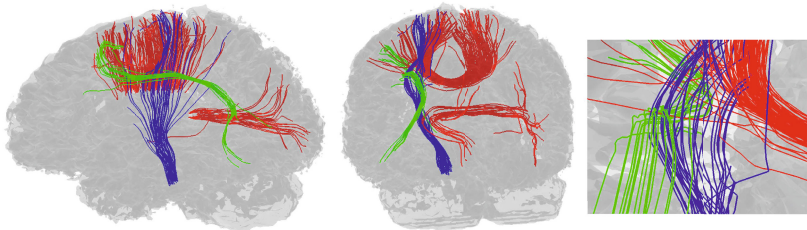
Fig. 5. Tractography from the corticospinal fiber bundle (blue), parts of the corpus callosum fiber bundles (red), and the superior longitudinal fasciculus (green). The figure on the right is a zoom on the crossing region between the three bundles.

Real data : We present results on a in vivo human cerebral dataset in Fig. 5. This figure shows the corticospinal fiber bundle (CST, in blue), parts of the corpus callosum fiber bundles (CC, in red), and the superior longitudinal fasciculus (green, SLF). These bundles agree with anatomical atlases of the white matter and our EAP-tract method allows a complete distinction between these three white matter fiber bundles.The zoom (Fig. 5 on the right) shows the crossing region between the CST, the CC and the SLF. Again, the EAP-tract method well resolve this crossing area.

6   Conclusions

In this paper, we proposed and motivated the use of the Ensemble Average Propagator for tractography in dMRI. To the best of our knowledge, it is the first attempt to use the directional information of the EAP at different radii. We illustrated the great potential of the EAP-tract method on synthetic, phantom and real data. These experiments showed that our method is especially efficient to resolve regions where fiber bundles are crossing, and still well handle other fiber bundle configurations such as U-shape and kissing. We also presented numerous examples showing the advantage of the EAP-tract method over the common ODF-tract method. Furthermore, by tracking from seeds in the CC, CST, and SLF, we were able to correctly trace through regions where all three fiber bundles cross.

References

1. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the ODF in single and multiple shell q-ball imaging within constant solid angle. Magn. Reson. Med. 64(2), 554–566 (2010)
2. Assemlal, H., Tschumperl, D., Brun, L.: Efficient and robust computation of pdf features from diffusion mr signal. Medical Image Analysis 13(5), 715–729 (2009)
3. Basser, P.J., Mattiello, J., Le Bihan, D.: Mr diffusion tensor spectroscopy and imaging. Biophysical Journal 66(1), 259–267 (1994)
4. Cheng, J., Ghosh, A., Jiang, T., Deriche, R.: Model-Free and Analytical EAP Reconstruction via Spherical Polar Fourier Diffusion MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 590–597. Springer, Heidelberg (2010)
5. Descoteaux, M., Deriche, R., Knosche, T.R., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Transactions in Medical Imaging 28(2), 269–286 (2009)
6. Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.F., Poupon, C.: Quantitative analysis of 10 tractography algorithms on a realistic diffusion MR phantom. Neuroimage 56(1), 220–234 (2011)
7. Ozarslan, E., Koay, C., Shepherd, T., Blackband, S., Basser, P.: Simple harmonic oscillator based reconstruction and estimation for three-dimensional q-space mri. In: ISMRM, p. 1396 (2009)
8. Behrens, T.E.J., Johansen-Berg, H., Woolrich, S.J.M.F.S.R., Probabilistic, M.W.: Probabilistic diffusion tractography with multiple fibre orientations. what can we gain? NeuroImage 34(1), 144–155 (2007)
9. Tristan-Vega, A., Westin, S., Aja-Fernandez, S.: Estimation of fiber orientation probability density functions in high angular resolution diffusion imaging. NeuroImage 47(2), 638–650 (2009)
10. Tuch, D.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)

A 4D Statistical Shape Model for Automated Segmentation of Lungs with Large Tumors

Matthias Wilms, Jan Ehrhardt, and Heinz Handels

Institute of Medical Informatics, University of Lübeck, Lübeck, Germany
wilms@imi.uni-luebeck.de

Abstract. Segmentation of lungs with large tumors is a challenging and time-consuming task, especially for 4D CT data sets used in radiation therapy. Existing lung segmentation methods are ineffective in these cases, because they are either not able to deal with large tumors and/or process every 3D image independently neglecting temporal information. In this paper, we present a approach for model-based 4D segmentation of lungs with large tumors in 4D CT data sets. In our approach, a 4D statistical shape model that accounts for inter- and intra-patient variability is fitted to the 4D image sequence, and the segmentation result is refined by a 4D graph-based optimal surface finding. The approach is evaluated using 10 4D CT data sets of lung tumor patients. The segmentation results are compared with a standard intensity-based approach and a 3D version of the presented model-based segmentation method. The intensity-based approach shows a better performance for normal lungs, however, fails in presence of large lung tumors. Although overall performance of 3D and 4D model-based segmentation is similar, the results indicate improved temporal coherence and improved robustness with respect to the segmentation parameters for the 4D model-based segmentation.

1 Introduction

Breathing-induced tumor motion represents a major challenge in radiation therapy of lung cancer. Patient-specific information about the respiratory dynamics, estimated by using spatio-temporal 4D CT data sets and non-linear image registration techniques, can therefore help to optimize the treatment planning and the delivery process [1]. To confine motion estimation and subsequent analysis steps (e.g., tumor detection and tracking) to the lungs, segmentations of the lung tissue in all 3D images of a 4D sequence are needed.

Many approaches dealing with the automatic segmentation of healthy lungs in 3D CT images have been proposed (e.g., [2,3]). These methods mainly take advantage of the large density difference between the air-filled lungs and surrounding tissue and therefore frequently fail to include areas of high density abnormalities (e.g., tumors) into the segmentation. As a consequence, several groups have suggested algorithms specifically designed to handle CT images of pathological lungs by incorporating prior knowledge to guide the segmentation process. For example, Sluimer et al. [4] and van Rikxoort et al. [5] employed

atlas-based techniques for the segmentation of lungs with arbitrary pathologic abnormalities. They were able to significantly increase segmentation accuracy, but the required non-linear atlas-to-image registration process was very time consuming. Sun et al. presented a time-efficient method especially aimed to segment lungs with large tumors using a robust active shape model [6]. Their results are very promising, but their approach, like other approaches, can only handle single 3D images. In the case of 4D data sets, methods processing all individual 3D images separately ignore the temporal information included in spatio-temporal 4D data: We hypothesize that including temporal information improves robustness of the segmentation process and temporal consistency of the results.

The aim of this work is the temporally consistent segmentation of lungs with large tumors in 4D CT data sets. Our method is based on the work of Perperidis et al. [7], who used a 4D statistical shape model (4D-SSM) for 4D cardiac image segmentation. This 4D-SSM accounts for both changes of the organ shape caused by inter-patient variability and shape changes due to cardiac (or in our case, respiratory) dynamics (intra-patient variability). We present a novel 4D fitting algorithm for this 4D-SSM and refine the segmentation results by using 4D graph-based optimal surface finding. In an evaluation, the results of our 4D approach are compared to segmentations obtained by using standard 3D-SSMs, as well as results of a standard intensity-based lung segmentation algorithm.

2   Method

First, we generate a 4D-SSM for each lung based on N_p segmented 4D CT image sequences of different patients (Section 2.1). Each image sequence is assumed to consist of N_j 3D image volumes $I_{p,j} : \Omega \to \mathbb{R}$ $(\Omega \subset \mathbb{R}^3)$, reconstructed at corresponding phases j of the breathing cycle, e.g. end-expiration (EE), mid-inspiration (MI), end-inspiration (EI), and mid-expiration (ME). For segmentation, the generated model is simultaneously adapted to all 3D images of an unseen 4D CT data set (Section 2.2). Afterwards, all segmentations are refined using a graph-based post-processing method (Section 2.3).

2.1   Building a 4D Statistical Shape Model

The first step in building a statistical shape model (SSM) based on a training set of $N = N_p N_j$ complete lung shapes obtained from segmented 4D CT image sequences is to establish correspondence between all shapes. This is achieved by propagating M pseudo-landmarks from an automatically landmarked atlas to all other shapes of the training set. After the generation and landmarking (by means of a surface triangulation and curvature-based mesh simplification) of an average lung shape atlas, landmark propagation is done by using non-linear transformations obtained from atlas-patient and intra-patient registrations of the images performed with a non-linear diffeomorphic registration method [8].

Let $\{\boldsymbol{q}_{p,j} \in \mathbb{R}^{3M} | p = 1, \ldots, N_p; j = 1, \ldots, N_j\}$ denote the set of N aligned training shapes. Each shape vector $\boldsymbol{q}_{p,j} = [\boldsymbol{p}_{p,j,1}^T, \ldots, \boldsymbol{p}_{p,j,M}^T]^T$ consists of a

concatenation of M landmarks $\boldsymbol{p}_{p,j,k} \in \Omega$. Principal component analysis (PCA) performed on this shape vectors yields a common 3D-SSM

$$S_{3D}(\boldsymbol{b}, \varphi) = \varphi(\overline{\boldsymbol{q}} + \boldsymbol{P}\boldsymbol{b}) \quad \text{,with} \quad \overline{\boldsymbol{q}} = \frac{1}{N} \sum_{p=1}^{N_p} \sum_{j=1}^{N_j} \boldsymbol{q}_{p,j} \ , \tag{1}$$

where $\boldsymbol{P}$ denotes a matrix whose columns are orthonormal eigenvectors $\boldsymbol{e}_i$ of covariance matrix $\boldsymbol{C} = 1/N \sum_{p=1}^{N_p} \sum_{j=1}^{N_j} (\boldsymbol{q}_{p,j} - \overline{\boldsymbol{q}})(\boldsymbol{q}_{p,j} - \overline{\boldsymbol{q}})^T$ with eigenvalues $\lambda_i < \lambda_{i+1}$ [9]. The model is parameterized by shape parameter vector $\boldsymbol{b}$ and similarity transformation φ. As a consequence of using all shapes of all patients to estimate $\boldsymbol{C}$, the resulting eigenmodes explain both inter- and intra-patient variability. This leads to N_j different and independent sets of model parameters $\{\boldsymbol{b}, \varphi\}$ to describe a lung's shape during a breathing cycle, making the model unsuitable for 4D segmentation. We therefore propose the application of a so-called 4D-SSM based on the work of Perperidis et al. [7] and defined by

$$S_{4D}(\boldsymbol{b}_{inter}, \boldsymbol{b}_{intra}^j, \varphi) = \varphi(\overline{\boldsymbol{q}} + \boldsymbol{P}_{inter}\boldsymbol{b}_{inter} + \boldsymbol{P}_{intra}\boldsymbol{b}_{intra}^j) \ , \tag{2}$$

which describes a patient's lung shape as a combination of a fixed patient-specific part (given by $\boldsymbol{b}_{inter}$ and φ) and a varying part depending on the breathing phase j (weighted by $\boldsymbol{b}_{intra}^j$). $\boldsymbol{P}_{inter}$ and $\boldsymbol{P}_{intra}$ denote matrices of orthonormal eigenvectors defining subspaces of $\mathbb{R}^{3M}$ accounting for inter- and intra-patient variability, respectively. Separate PCAs performed on the covariance matrices $\boldsymbol{C}_{inter}$ and $\boldsymbol{C}_{intra}$ yield the eigenvectors defining both subspaces. $\boldsymbol{C}_{inter}$ explains the variability across the different patients and is given by

$$\boldsymbol{C}_{inter} = \frac{1}{N_p} \sum_{p=1}^{N_p} (\overline{\boldsymbol{q}}_p - \overline{\boldsymbol{q}})(\overline{\boldsymbol{q}}_p - \overline{\boldsymbol{q}})^T \quad \text{,with} \quad \overline{\boldsymbol{q}}_p = \frac{1}{N_j} \sum_{j=1}^{N_j} \boldsymbol{q}_{p,j} \ . \tag{3}$$

Accordingly, $\boldsymbol{C}_{intra}$ describes the shape differences across the respiratory cycle as deviations from the mean shape of each patient p:

$$\boldsymbol{C}_{intra} = \frac{1}{N_p N_j} \sum_{p=1}^{N_p} \sum_{j=1}^{N_j} (\boldsymbol{q}_{p,j} - \overline{\boldsymbol{q}}_p)(\boldsymbol{q}_{p,j} - \overline{\boldsymbol{q}}_p)^T \ . \tag{4}$$

2.2 Fitting the Model to an Image Sequence

The simultaneous lung segmentation in all N_j 3D images I_j, $j \in \{1 \ldots, N_j\}$, of a 4D CT data set using a 4D-SSM (eq. 2) consists of finding parameters $\{\tilde{\boldsymbol{b}}_{inter}, \tilde{\boldsymbol{b}}_{intra}^1, \ldots, \tilde{\boldsymbol{b}}_{intra}^{N_j}, \tilde{\varphi}\}$, such that $\tilde{\boldsymbol{r}}_j = S_{4D}(\tilde{\boldsymbol{b}}_{inter}, \tilde{\boldsymbol{b}}_{intra}^j, \varphi)$ is a good approximation of lung shape $\boldsymbol{r}_j \in \mathbb{R}^{3M}$ implicitly encoded in I_j. When using the common sum of squared distances between corresponding model landmarks and image points as a measure, this can be formulated as the optimization problem

$$\sum_{j=1}^{N_j} \|\boldsymbol{r}_j - S_{4D}(\boldsymbol{b}_{inter}, \boldsymbol{b}_{intra}^j, \varphi)\|^2 \longrightarrow \min \ , \tag{5}$$

which we divide into

$$(\tilde{\boldsymbol{b}}_{inter}, \tilde{\varphi}) = \underset{\boldsymbol{b}_{inter}, \varphi}{\operatorname{argmin}} \| \frac{1}{N_j} \sum_{j=1}^{N_j} \boldsymbol{r}_j - S_{4D}(\boldsymbol{b}_{inter}, 0, \varphi) \|^2 \text{ and} \tag{6}$$

$$(\tilde{\boldsymbol{b}}_{intra}^j) = \underset{\boldsymbol{b}_{intra}^j}{\operatorname{argmin}} \| \boldsymbol{r}_j - S_{4D}(\tilde{\boldsymbol{b}}_{inter}, \boldsymbol{b}_{intra}^j, \tilde{\varphi}) \|^2 . \tag{7}$$

Eq. (6) is motivated by the assumption that patient-specific shape properties independent of the breathing motion can be described by the patient's mean shape (eq. (3)). In contrast to [7], both problems are minimized by an alternating iterative optimization scheme based on the active shape model (ASM) algorithm [9] without a heuristic pre-initialization of the intra-patient parameters $\boldsymbol{b}_{intra}^j$:

1. Initial placing of mean shape $S_{4D}(0, 0, \varphi)$. The initial φ is determined by a heuristic based on the detected bronchial tree [10].
2. For all $j \in \{1 \ldots, N_j\}$: Displace each landmark $\boldsymbol{p}_{j,k}$ of model instance $S_{4D}(\tilde{\boldsymbol{b}}_{inter}, \tilde{\boldsymbol{b}}_{intra}^j, \tilde{\varphi})$ to better match the corresponding lung surface in I_j. The displaced landmarks $\hat{\boldsymbol{p}}_{j,k}$ form a candidate shape vector $\boldsymbol{r}_j \in \mathbb{R}^{3M}$.
3. The mean candidate shape vector $\overline{\boldsymbol{r}} = 1/N_j \sum_{j=1}^{N_j} \boldsymbol{r}_j$ is used to determine $\tilde{\varphi}$ and $\tilde{\boldsymbol{b}}_{inter}$ (see eq. (6)) by the minimization of $\| \overline{\boldsymbol{r}} - \varphi(\overline{\boldsymbol{q}}) \|^2$, and $\tilde{\boldsymbol{b}}_{inter} = \boldsymbol{P}_{inter}^T(\tilde{\varphi}^{-1}(\overline{\boldsymbol{r}}) - \overline{\boldsymbol{q}})$.
4. Repeat step 2. Each new $\boldsymbol{r}_j$ is used to determine a corresponding breathing-related $\tilde{\boldsymbol{b}}_{intra}^j$. Solving (7) yields $\tilde{\boldsymbol{b}}_{intra}^j = \boldsymbol{P}_{intra}^T(\tilde{\varphi}^{-1}(\boldsymbol{r}_j) - \overline{\boldsymbol{q}} - \boldsymbol{P}_{inter}\tilde{\boldsymbol{b}}_{inter})$.
5. Steps 2-4 are repeated until convergence.

A displaced landmark's position $\hat{\boldsymbol{p}}_{j,k} = \boldsymbol{p}_{j,k} + \boldsymbol{n}_{j,k} \delta s_{j,k}$ is determined by choosing the optimal position

$$s_{j,k} = \underset{l=-L,\ldots,+L}{\operatorname{argmin}} F_{j,k}(\boldsymbol{p}_{j,k} + \boldsymbol{n}_{j,k}\delta l) \tag{8}$$

on a sampled 1D intensity profile of $2L+1$ points at intervals of δ along the unit surface normal $\boldsymbol{n}_{j,k}$. Each sampling point is evaluated by

$$F_{j,k}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } I_j(\boldsymbol{x}) > -100\,\text{HU} \\ 1 - \frac{\max\{0, \boldsymbol{n}_{j,k}^T \nabla I_j(\boldsymbol{x})\}}{g_{max,j}} & \text{otherwise} \end{cases} \tag{9}$$

where $g_{max,j}$ is the maximum gradient magnitude in I_j. The value $F_{j,k}(\boldsymbol{x})$ is inversely related to the likelihood that $\boldsymbol{x}$ is a point on the lung's surface.

Success of the model fitting largely depends on the selected displaced landmarks. If many landmarks are detected at positions representing transitions from healthy lung parenchyma to tumor tissue, the outlined least squares approach will fail to recover the true lung shape. We try to avoid this by assuming an initial position close to the lung shape, and therefore use only short profiles.

2.3 Optimal Surface Finding

The final part of our segmentation approach aims at integrating patient-specific shape variations, not described by 4D-SSMs built from usually small training sets. Therefore, the N_j shapes $\{\tilde{\boldsymbol{r}}_j\}$ resulting from the model fitting step are deformed to better match the image data, while maintaining the established spatial and temporal consistency (optimal surface finding). Let $\tilde{R}_j = (V_j, E_j)$ denote the triangulated surface mesh of shape vector $\tilde{\boldsymbol{r}}_j$ with vertex set V_j and edge set E_j. The deformation of all surfaces is, as in Sec. 2.2, achieved by displacing every vertex $\boldsymbol{p}_{j,k} \in V_j$ along its surface normal to a position $\hat{\boldsymbol{p}}_{j,k} = \boldsymbol{p}_{j,k} + \boldsymbol{n}_{j,k}\delta s_{j,k}$. In contrast to the model fitting step, where all displacements $\{s_{j,k} \in [-L.. + L]\}$ are detected independently, globally optimal solutions are needed to preserve the consistencies mentioned above. The task of determining spatially consistent refined segmentations can be defined as the optimization problem

$$\sum_{j=1}^{N_j} \sum_{k=1}^{M} F_{j,k}(\boldsymbol{p}_{j,k} + \boldsymbol{n}_{j,k}\delta s_{j,k}) + \sum_{j=1}^{N_j} \sum_{[\boldsymbol{p}_{j,k}, \boldsymbol{p}_{j,m}] \in E_j} a|s_{j,k} - s_{j,m}| \xrightarrow{\{s_{j,k}\}} \min \quad (10)$$

$$\text{subject to} \quad \forall j \in \{1, \ldots, N_j\} \, \forall [\boldsymbol{p}_{j,k}, \boldsymbol{p}_{j,m}] \in E_j : |s_{j,k} - s_{j,m}| \leq \Delta_{3D}$$

where Δ_{3D} is the parameter of a hard smoothness constraint specifying how many steps adjacent vertices are allowed to shift against each other on their sampled profiles, while constant a penalizes every shift (soft smoothness constraint). Both constraints aim to prevent large deviations from the prior shapes $\{\tilde{\boldsymbol{r}}_j\}$. A globally optimal solution of eq. (10) can be obtained by computing the minimum-closed-set of a directed arc-weighted graph build from all sampled profiles using a max-flow/min-cut algorithm [11]. For our 4D segmentation approach, we also try to maintain the established temporal consistency by incorporating an additional hard smoothness constraint into eq. (10), where Δ_{4D} limits the shifting of temporally corresponding vertices:

$$\forall m \in \{1, \ldots, N_j\} \, \forall n \in \{m, \ldots, N_j\} \, \forall k \in \{1, \ldots, M\} : |s_{m,k} - s_{n,k}| \leq \Delta_{4D} \ .$$

3 Experiments and Results

12 4D CT data sets of healthy lungs with $N_j = 10$ phases are used to build 3D-SSMs (eq. (1)) and 4D-SSMs (eq. (2)) for left and right lung separately ($N = 120$ shapes with $M \approx 2000$ landmarks). The most significant *inter*-patient eigenmode of the 4D-SSM describes lung shapes from high and thin to low and broad, while the most significant *intra*-patient mode explains most of the breathing-related volume changes.

10 4D CT data sets of lung cancer patients (each containing 7-14 3D images) are used for the evaluation. All images had a size of $512 \times 512 \times 126$-467 voxel with a voxel size between $0.94 \times 0.94 \times 1.5 \, \text{mm}^3$ and $0.97 \times 0.97 \times 3.0 \, \text{mm}^3$. Out of the 20 different lungs, 8 were without abnormalities (normal lungs), 7 contained small tumors $< 13 \, \text{cm}^3$, and 5 contained large tumors $> 13 \, \text{cm}^3$ adhering to non-lung structures. Manual segmentations were available for the breathing phases

Table 1. Performance comparison between 4D-Seg, 3D-Seg, standard intensity-based (Conv), and combined (Comb) segmentation approaches. Results averaged over all processed lungs of each group (normal & small tumors, large tumors), given as $\mu \pm \sigma$. Error metrics: Jaccard coefficient $J(A, B)$, symmetric mean surface distance $\overline{D}(A, B)$, and the symmetric Hausdorff distance $H(A, B)$. A and B are the automatically estimated lung region and the corresponding manual segmentation serving as ground truth.

Measures/Methods	4D-Seg	3D-Seg	Conv	Comb
	Normal lungs/Lungs with small tumors $< 13\,\mathrm{cm}^3$			
$J(A, B)$	0.92 ± 0.03	0.92 ± 0.03	0.95 ± 0.02	0.95 ± 0.02
$\overline{D}(A, B)$ [mm]	1.32 ± 0.48	1.30 ± 0.51	0.91 ± 0.52	0.85 ± 0.31
$H(A, B)$ [mm]	25.06 ± 9.53	24.89 ± 8.67	22.68 ± 8.98	22.98 ± 8.86
	Lungs with large tumors $> 13\,\mathrm{cm}^3$			
$J(A, B)$	0.92 ± 0.03	0.92 ± 0.03	0.89 ± 0.04	0.95 ± 0.02
$\overline{D}(A, B)$ [mm]	1.45 ± 0.49	1.46 ± 0.52	2.12 ± 1.26	0.91 ± 0.30
$H(A, B)$ [mm]	21.57 ± 6.02	21.69 ± 5.93	36.65 ± 11.03	19.64 ± 3.63

(a) 4D-Seg   (b) 3D-Seg   (c) Conv   (d) Comb
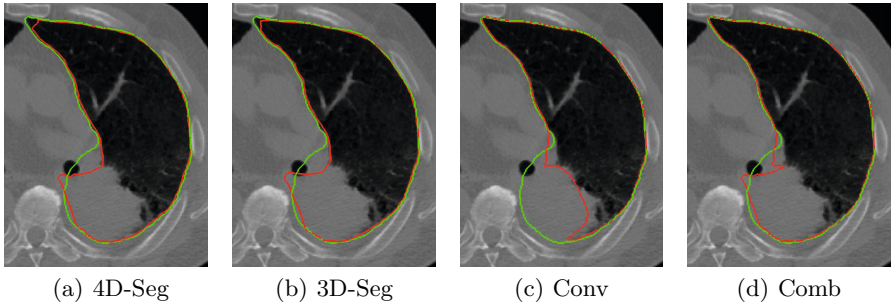
Fig. 1. Segmentation results for a right lung with a tumor adhering to the chest wall. Results generated with different segmentation methods. Depicted is an axial slice of one breathing phase (EI) of a 4D CT image sequence. Results of the automatic methods are displayed in red, the manually obtained segmentation in green.

(a) EE   (b) MI   (c) EI   (d) ME
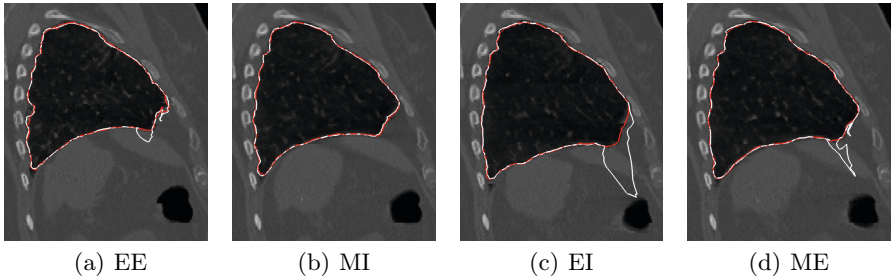
Fig. 2. Left lung segmentation results for 4 different respiratory phases of a 4D CT image sequence obtained by employing 3D-Seg (white contour) and 4D-Seg (red contour) method. Results of 3D-Seg are partially influenced by the gas-filled bowel resulting in temporally inconsistent segmentations. The errors in (c) and (d) are mainly introduced in the 3D-SSM fitting stage. Results of 4D-Seg were computed using all 10 phases of the 4D data set.

of EE, MI, EI, and ME. In total, 80 segmentations (normal and small tumors: 60, large tumors: 20) were used as ground truth for the evaluation. The accuracy of the proposed model-based 4D segmentation approach (4D-Seg) was compared to three other approaches: 1) standard intensity-based lung segmentation similar to [3] (Conv), 2) standard 3D-SSMs (see eq. (1)) and ASM fitting with spatially consistent optimal surface finding (3D-Seg), and 3) combined results of 4D and standard segmentation (Comb=4D-Seg $\cup$ Conv). Three error metrics were utilized for the comparison: Jaccard coefficient, symmetric mean surface distance, and the symmetric Hausdorff distance. The parameters of method 4D-Seg were fixed for all test cases ($L = 20$, $\delta = 1.0$ mm, $\Delta_{3D} = \Delta_{4D} = 10$, $a = 0.01$). For the 3D-Seg method, we were unable to determine common parameter values suitable for all test cases ($L = 15$-30, $\delta = 1.0$ mm, $\Delta_{3D} = 10$, $a = 0.01$-0.015).

Table 1 summarizes the quantitative results of our evaluation grouped into two categories (normal and small tumors, large tumors). It can be observed that standard segmentation (Conv) leads to better results for normal lungs and lungs with small tumors than the model-based approaches 4D-Seg and 3D-Seg. As shown in Fig. 1, in the case of lungs with large tumors adhering to non-lung structures, Conv is outperformed by 4D-Seg and 3D-Seg, which give nearly equivalent overall accuracy. The small differences between them reported in Tab. 1 are not statistically significant (paired t-test, $p < 0.05$). However, despite that, Fig. 2 depicts an exemplary case where only the 3D approach is partially influenced by the gas-filled bowel resulting in temporally inconsistent segmentations. Using 3D-Seg with a single set of parameters for all patients ($L = 30$, $a = 0.01$) leads to two additional cases with temporal inconsistencies.

Due to the globally chosen smoothness constraints and the small training set used for model building, both model-based approaches are prone to cause over- and under-segmentation in higly curved areas of the lungs. Therefore, combining the results of 4D and standard segmentation (Comb) yields the best accuracy for lungs with large tumors, because under-segmentation is considerably reduced.

4   Conclusion

In this paper, we present an automatic model-based method to simultaneously segment the lungs in all 3D images of 4D CT data sets of lung cancer patients. This method combines the fitting of a 4D-SSM with a 4D graph-based refinement step, taking into account spatio-temporal consistency. An intrinsic characteristic of the 4D-SSM is that differences within a 4D sequence are restricted to intra-patient variations and therefore temporal consistency is achieved without explicit temporal regularization. In contrast to [7], our novel 4D model fitting algorithm works without a heuristic pre-initilization of the intra-patient variation parameters. Thus, no assumptions about number and ordering of the breathing phases to be segmented are needed. Furthermore, the 4D-SSM can also be used to segment breathing phases not included in the model's training data set. Experimental results demonstrate the potential of the proposed 4D model-based approach, which performs better than standard intensity-based segmentation in the presence of

large tumors adhering to non-lung structures. Under-segmentations in highly curved areas of the lungs can be reduced by combining 4D and conventional segmentation algorithms. While the mean overall accuracy of 3D and 4D model-based segmentation is nearly identical, improvements in temporal coherence and robustness with respect to the segmentation parameters can be achieved by using the 4D approach. Summing up, there is no reason to use a 3D approach for lung segmentation in 4D data sets in the presence of large tumors. This is also supported by the nearly identical running times of both model-based methods ($\approx$40 min. for a 4D data set with 14 phases). In future work, the training and test data sets will be significantly enlarged, and we intend to improve the robustness of the least-squares model fitting by adding an outlier detection step.

References

1. Keall, P.J., Mageras, G., Balter, J.M., et al.: The management of respiratory motion in radiation oncology report of AAPM task group 76. Med. Phys. 33(10), 3874–3900 (2006)
2. Armato, S.G., Sensakovic, W.F.: Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis. Acad. Radiol. 11(9), 1011–1021 (2004)
3. Hu, S., Hoffman, E., Reinhardt, J.: Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans. Med. Imaging 20(6), 490–498 (2001)
4. Sluimer, I., Prokop, M., van Ginneken, B.: Towards automated segmentation of the pathological lung in CT. IEEE Trans. Med. Imaging 24(8), 1025–1038 (2005)
5. van Rikxoort, E., de Hoop, B., Viergever, M., Prokop, M., van Ginneken, B.: Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. Med. Phys. 36(7), 2934–2947 (2009)
6. Sun, S., Bauer, C., Beichel, R.: Automated 3-d segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach. IEEE Trans. Med. Imaging 31(2), 449–460 (2012)
7. Perperidis, D., Mohiaddin, R., Edwards, P., Rueckert, D.: Segmentation of cardiac mr and ct image sequences using model-based registration of a 4d statistical model. In: Pluim, J.P.W., Reinhardt, J.M. (eds.) Proc. SPIE Medical Imaging 2007: Image Processing. SPIE, vol. 6512, p. 65121D (2007)
8. Ehrhardt, J., Werner, R., Schmidt-Richberg, A., Handels, H.: Statistical modeling of 4D respiratory lung motion using diffeomorphic image registration. IEEE Trans. Med. Imaging 30(2), 251–265 (2011)
9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models their training and application. CVIU 61, 38–59 (1995)
10. Mori, K., Hasegawa, J., Toriwaki, J., Anno, H., Katada, K.: Recognition of bronchus in three-dimensional X-ray CT images with application to virtualized bronchoscopy system. In: Pluim, J.P.W., Reinhardt, J.M. (eds.) Proc. 13th Int. Conf. Pattern Recognition, vol. 3, pp. 528–532. IEEE Computer Society (1996)
11. Song, Q., Wu, X., Liu, Y., Smith, M., Buatti, J., Sonka, M.: Optimal Graph Search Segmentation Using Arc-Weighted Graph for Simultaneous Surface Detection of Bladder and Prostate. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 827–835. Springer, Heidelberg (2009)

Closed-Form Relaxation for MRF-MAP Tissue Classification Using Discrete Laplace Equations

Alexis Roche

Siemens Research, CIBM, Lausanne, Switzerland
`alexis.roche@epfl.ch`

Abstract. While Markov random fields are very popular segmentation models in medical image processing, the associated maximum a posteriori (MAP) estimation problem is usually solved using iterative methods that are prone to local maxima. We show that a variant of the random walker algorithm can be seen as a relaxation method for the MAP problem under the Potts model. The key advantage of this technique is that it boils down to a sparse linear system with a uniquely defined explicit solution. Our experiments further demonstrate that the resulting MAP approximation can be used to improve the classical mean-field algorithm in terms of MAP estimation quality.

1 Introduction

Many image segmentation problems can be conveniently formulated using Markov random fields (MRF), however the associated task of computing the maximum a posteriori (MAP) segmentation is combinatorial NP-hard. Early MRF-MAP tracking methods include the ICM algorithm [1] known to be fast but highly prone to local maxima, and simulated annealing [2] which may be hopelessly slow in practice. Over the past two decades, several approaches have been proposed to work around these limitations.

One such approach, which stems from classical optimization theory, is relaxation. The basic idea is to substitute the combinatorial optimization problem with a continuous one which, in image segmentation context, involves extending the MAP search to the space of probabilistic assignments from voxels to classes. An approximation to the MAP is found by binarizing the optimal such assignment. Relaxation for MRF-MAP has been implemented using convex programing [3,4,5], which guarantees a unique solution but tends to be computationally expensive as it relies on constrained optimization.

Meanwhile, message-passing algorithms have emerged from the machine learning community for inference on probabilistic graphical models [6]. In particular, the variational expectation-maximization (VEM) algorithm, also known as mean-field algorithm, has long been used in brain imaging, though sometimes through ad-hoc variants [7,8,9,10]. Other message-passing schemes used in computer vision include belief propagation and tree-reweighted message-passing [11]. Message-passing can be viewed as a special kind of relaxation for MRF-MAP,

which is rather fast as it does do not require handling explicit inequality constraints, but is initialization-dependent since the underlying objective function is typically non-convex.

Another research trend has been to apply methods from deterministic graph theory to image segmentation. Nowadays, graph cuts [12] are widely regarded as the most robust methods for MRF-MAP [11]. In the binary segmentation case where MRF-MAP amounts to a max-flow min-cut problem, they are guaranteed to find the global maximum if unique. However, conventional graph cut methods such as the expansion and swap algorithms iterate over labels or pairs of labels when more than two labels are involved, and thus become prone to local convergence in addition to being slower.

A method that closely relates with graph cuts is the random walker (RW) algorithm of Grady [13,14], which was previously exposed in a different and less general form by Marroquin *et al* [15]. In this work, we show how to re-tune RW to yield a powerful MRF-MAP relaxation method that boils down to solving a sparse linear system. Although RW stems from a variational problem similar to the min-cut in the two-label case [16], its deep connections with MRF-based segmentation have been somewhat overlooked so far. We further advocate a method that combines the proposed relaxation with the traditional VEM algorithm.

2   MRF-MAP Segmentation

The MRF-MAP problem under the Potts model for labeling an image Y in K classes amounts to minimizing the following energy [17]:

$$L(\delta, \theta) = -\sum_i \delta_i^\top \log \ell_i(\theta) + \beta \sum_{i,j} w_{ij}(1 - \delta_i^\top \delta_j), \qquad (1)$$

where $\delta = (\delta_1, \delta_2, \ldots)$ is a collection of "delta-distributions", that is, for each voxel i, δ_i is a K-dimensional vector with a single non-zero component $\delta_{ik} = 1$ corresponding to the voxel label. The weights w_{ij} encode spatial interactions between voxels and are usually symmetric, equal to one if voxels i and j are neighbors according to a given discrete topology, and zero otherwise. The first term in the right hand side involves the likelihood $\ell_{ik}(\theta) = p(y_i|k, \theta)$ of the labels at voxel i, where θ is a nuisance parameter vector to be estimated. Under the usual Gaussian noise model, $p(y_i|k, \theta) = N(y_i; \mu_k, \sigma_k)$ and $\theta = (\mu_1, \sigma_1, \ldots, \mu_K, \sigma_K)$ is the concatenation of mean intensities and standard deviations over classes. Note, however, that the likelihood can be substituted with any external field without changing the analysis that follows.

2.1   Free Energy Relaxation

A known relaxation method to approximate the minimization of (1) is to minimize the so-called free energy function over arbitrary probability masses q_i,

$$\tilde{L}(q, \theta) = -\sum_i q_i^\top \log \ell_i(\theta) + \beta \sum_{i,j} w_{ij}(1 - q_i^\top q_j) + \sum_i q_i^\top \log q_i, \qquad (2)$$

which defines a continuous extension of (1) in the sense that $\tilde{L}(\delta, \theta)$ and $L(\delta, \theta)$ coincide on the space of delta-distributions. While the global minimization of (2) is intractable, one may resort to a greedy approach based on an alternate minimization along the q_i's, yielding explicit updates [10]:

$$q_i \propto \ell_i(\theta)e^{2\beta \sum_j w_{ij}q_j}$$

Applying this equation iteratively by cycling through the voxels corresponds to the E-step of a VEM algorithm and is guaranteed, under broad conditions, to converge to a local minimum of free energy. In the VEM algorithm, probability updates are interleaved with minimizations along θ to concurrently refine intensity parameters. The VEM algorithm is nevertheless dependent on starting values for both q and θ, and there is no theoretical warranty as to the accuracy of the resulting MAP approximation even at fixed θ.

2.2 Laplace Relaxation

We now describe another MAP relaxation approach that yields a convex problem unlike (2). Let us start with defining a surrogate MAP energy function:

$$L_s(\delta, \theta) = \sum_i (1 - \delta_i^\top \pi_i(\theta)) + \beta \sum_{i,j} w_{ij}(1 - \delta_i^\top \delta_j) - \sum_i \log z_i(\theta),$$

where $\pi_i(\theta) = \ell_i(\theta)/z_i(\theta)$ is the likelihood at voxel i normalized to unit sum and $z_i(\theta)$ is the associated partition function. Using the inequality $\log(x) \le x - 1$, we see that $L_s(\delta, \theta) \le L(\delta, \theta)$ for any delta-distribution, with equality iff π_i is a delta-distribution and $\delta_i = \pi_i$ at each voxel. Moreover, we have:

$$\|\delta\|^2 = 1, \qquad 1 - \delta^\top v = \frac{1}{2}\|\delta - v\|^2 + \frac{1}{2} - \frac{1}{2}\|v\|^2,$$

for any delta-distribution δ and vector v. Therefore, the following function defines a continuous extension of L_s over arbitrary distributions:

$$\tilde{L}_s(q, \theta) = \frac{1}{2}\sum_i \|q_i - \pi_i(\theta)\|^2 + \frac{\beta}{2}\sum_{i,j} w_{ij}\|q_i - q_j\|^2 + C(\theta), \qquad (3)$$

where $C(\theta) = \sum_i(-\log z_i(\theta) + \frac{1}{2} - \frac{1}{2}\|\pi_i(\theta)\|^2)$. Clearly, $\tilde{L}_s$ is quadratic and strictly convex in q. Minimizing it at fixed θ yields the first-order condition:

$$\forall k, \quad (\mathbb{I} + \lambda L)Q_k = \Pi_k, \qquad \text{with} \quad \lambda = 2\beta, \qquad (4)$$

which is a set of sparse linear systems, where L is the Laplacian matrix of the image grid considered as a graph with weights w_{ij} and $\mathbb{I}$ is the identity matrix with size equal to the number of voxels. Q_k stands for the probability image associated with class k, i.e. $Q_{ki} = q_{ik}$, and Π_k similarly represents the normalized likelihood image for class k.

Equation (4) turns out to be a vector-valued discrete Laplace equation and is equivalent to the method proposed by Marroquin *et al* [15] as an approximation to the same MAP problem. The key property is that the unique solution is a probability map without the need to incorporate explicit equality or inequality constraints, which provides a massive computational advantage over other relaxation approaches [3,4,5]. In [15], the smoothing parameter λ was not related to the MRF parameter β and was tuned empirically. We showed here that setting $\lambda = 2\beta$ ensures that the surrogate energy L_s is uniformly upper bounded by the MAP objective (1), therefore (4) qualifies as a relaxation method for the MAP problem.

A generalization of (4) is the multilabel RW algorithm [13,14]. Our strategy should however be expected to differ significantly from the original RW in practice, since both the weights w_{ij} and the "diagonal matrix" (here, the identity) are chosen in different ways, independently from the data in our case.

Owing to the inequality $L_s(\delta, \theta) \leq L(\delta, \theta)$, the MAP may be bracketed using the Laplace relaxation solution $q_\star$ and its binarization $\delta_\star$:

$$\tilde{L}_s(q_\star) \leq \min_\delta L(\delta) \leq L(\delta_\star),$$

hence providing some confidence bounds on the MAP approximation. Note that such a lower bound is not available for the VEM output as it may not be a global minimizer of free energy (2).

Also, we shall note that there is no explicit solution to minimizing (3) with respect to θ, unlike the case of free energy. Therefore, Laplace relaxation does not come with a simple built-in method for intensity parameter estimation.

3   Experiments

This section compares both relaxation methods presented above in brain tissue classification. We used a subset of 248 brain MR T1-weighted images from the Alzheimer's Disease Neuroimaging Initiative database (ADNI, `adni.loni.ucla.edu`) acquired on both 1.5 Tesla and 3 Tesla scanners from different manufacturers, with voxel volume ranging from 1 to 1.9 mm^3. The dataset includes 163 healthy controls and 85 diagnosed AD patients (55% males, 45% females) with mean age 77 ± 7 years. As a pre-processing, the images were corrected for bias field using the N3 method [18] and skull-stripped by non-rigid registration with a template [19].

We here focus on further classifying the skull stripped data into cerebrospinal fluid (CSF), gray matter (GM), and white matter (WM). To this end, we used a 4-class Potts prior model using a 6-neighborhood system with two classes representing GM to account for the usually rather large intensity variations between cortical GM and deep GM in T1-weighted images. The spatial regularization parameter was set to $\beta = 0.5$ based on previous tests. No external field was incorporated to the model at this stage to avoid biasing tissue classification towards an atlas [20]. This model was found to yield high overlap with ground

truth segmentation on Brainweb data [21] using the conventional VEM algorithm (Jaccard indices larger than 0.88).

The VEM algorithm was mainly implemented in Python based on the Scientific Python package (`www.scipy.org`) with a subroutine in C for higher performance. Laplace relaxation was implemented in pure Python inspired by the random walker implementation by E. Gouillart (`github.com/emmanuelle`) using a smoothed aggregation solver [22]. The θ parameter supplied to each method was computed using a simple moment matching technique [10].

The computation time on a single processor Intel Core i7-975 CPU 3.33GHz was about 1.5 seconds per iteration for VEM (including update of θ), and about 15 seconds for Laplace relaxation. The VEM algorithm was run for 50 iterations, which achieved satisfactory convergence in all cases (relative variations of free energy lower than 2.5×10^{-4}), resulting in a total computation time of 75 seconds per image.
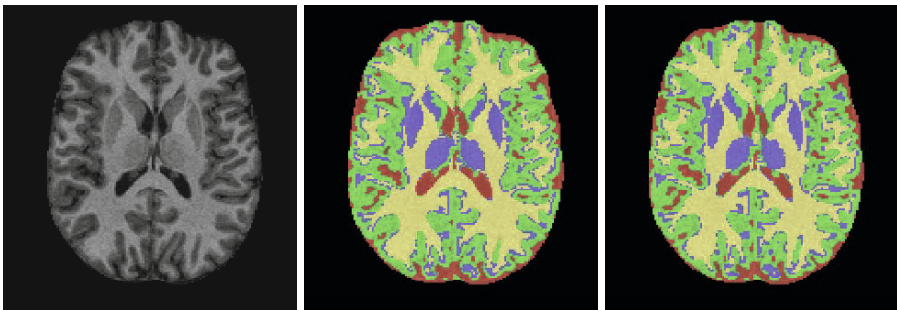
Fig. 1. Comparison of MAP estimates found by Laplace relaxation (middle) and the VEM algorithm (right) for a skull-stripped MR T1-weighted image (left). Label colors are red for CSF, blue and green for GM, and yellow for WM.

Figure 1 illustrates that MAP estimates found by Laplace relaxation generally look very similar from visual inspection to those provided by the VEM algorithm. To quantify this, we computed minimum Jaccard overlap coefficients,

$$J = \min_k \frac{|A_k \cap B_k|}{|A_k \cup B_k|},$$

where A_k and B_k denote the sets of voxel labeled as k in the respective classifications. Overlap coefficients ranged from 0.33 to 0.87 on the whole dataset with mean 0.64 and standard deviation 0.13. They were found from ANOVA to correlate negatively with voxel volume (p-value $< 10^{-10}$) and, to a lesser extent, with pathology (p-value $< 10^{-2}$), the agreement between both segmentation methods being higher for AD patients. Correlations with age and gender were not significant.

While the VEM algorithm was slower than Laplace relaxation, it converged to a solution of lower energy (1) in all of the 248 cases and was therefore more accurate at tracking the MAP despite being theoretically prone to local minima. To further investigate the benefit of Laplace relaxation, we tested a VEM variant, hereafter referred to as LR-VEM, where the class probability map q is initialized as the binarized solution of Laplace relaxation, as opposed to the standard initialization with a uniform distribution over labels.
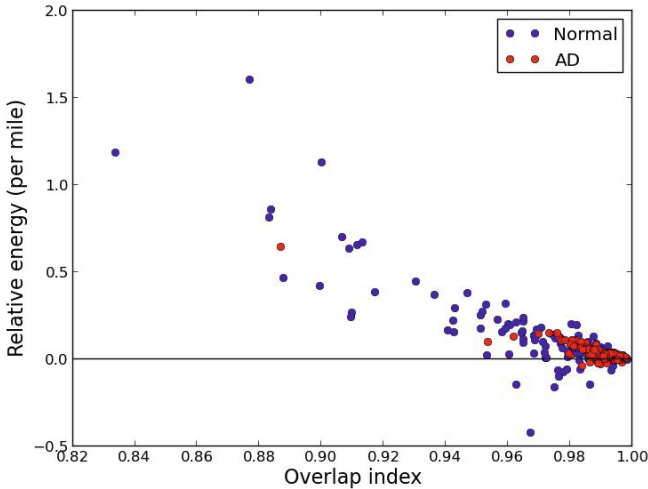
Fig. 2. Comparison between standard VEM and LR-VEM algorithms: plot of final relative MAP energy values against Jaccard indices for 248 ADNI subjects

Figure 2 plots a relative measure of MAP estimation quality of LR-VEM versus VEM, defined as $L_{final}^{VEM}/L_{final}^{LR\text{-}VEM} - 1$, where L_{final} denotes the energy level reached after 50 iterations, against the overlap indices computed between the respective corresponding MAP estimates. In 83.5% cases, LR-VEM achieved lower energy than VEM, while the converse happened in 16.5% cases. Segmentation results showed non-negligible differences in 10% cases as shown by overlap indices lower than 0.95. In all such cases, the MAP estimate from LR-VEM had the lower energy. Conversely, when VEM achieved lower energy than LR-VEM, the respective MAP estimates were almost identical. This provides some evidence that initialization with Laplace relaxation makes the VEM algorithm more robust in tracking the MAP.

ANOVA revealed that overlap indices correlate strongly with pathology (p-value $< 10^{-6}$) and age (p-value $< 10^{-5}$) in the sense that differences between VEM and LR-VEM are reduced for diseased or aged subjects. A slight negative correlation with voxel volume (p-value $< 10^{-2}$) was found in this case, and again no significant correlation with gender.

Moreover, the LR-VEM algorithm required an average of 7 ± 10 iterations less than VEM to achieve the same tolerance on free energy variations as in the last VEM iteration, meaning that the computational overhead of Laplace relaxation is compensated for by faster convergence.

4   Discussion

Laplace relaxation offers a fast alternative to the VEM algorithm for MRF-MAP classification that is independent from an initial label probability assignment. In our brain tissue classification experiments, Laplace relaxation produced results quite similar to the VEM algorithm (as shown by minimum Jaccard indices of 0.64 ± 0.13). The MAP classifications output by VEM were, however, more accurate. In a different scenario, Laplace relaxation can be used as an initialization step for the VEM algorithm, leading to noticeable improvements in MAP estimation in about 10% cases without significant overhead in computation time due to faster convergence.

We did not expect massive improvements in the whole-brain classification setting where the VEM algorithm has previously been reported to be robust. We anticipate that the effect of Laplace relaxation may be more substantial in segmentation applications that target specific anatomical structures since local volume or shape assessments are likely to be sensitive to small variations in tissue probability maps. The benefit of Laplace relaxation in brain morphometry is thus to be further investigated.

Laplace relaxation is currently applicable to a subclass of MRF models that includes extensions of the Potts model that involve non-stationary scalar-weighted interactions and addition of any external field. Future work will aim to extend the methodology to other MRF models for which iterative methods such as the VEM algorithm may have serious local convergence issues, in particular models that incorporate strong topological constraints via tissue-dependent interaction potentials.

References

1. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B 36(2), 192–236 (1974)
2. Geman, S., Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6(6), 721–741 (1984)
3. Ravikumar, P., Lafferty, J.: Quadratic Programming Relaxations for Metric Labeling and Markov Random Field MAP Estimation. In: International Conference on Machine Learning, pp. 737–744 (2006)

4. Cour, T., Shi, J.: Solving Markov Random Fields with Spectral Relaxation. Journal of Machine Learning Research 2, 75–82 (2007)
5. Kumar, M.P., Kolmogorov, V., Torr, P.H.: An Analysis of Convex Relaxations for MAP Estimation of Discrete MRFs. Journal of Machine Learning Research 10, 71–106 (2008)
6. Minka, T.P.: Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Research Ltd., Cambridge, UK (December 2005)
7. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging 18(10), 897–908 (1999)
8. Zhang, Y., Brady, J., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Transactions on Medical Imaging 20(1), 45–57 (2001)
9. Forbes, F., Fort, G.: Combining Monte Carlo and Mean-Field-Like Methods for Inference in Hidden Markov Random Fields. IEEE Transactions on Image Processing 16(3), 824–837 (2007)
10. Roche, A., Ribes, D., Bach-Cuadra, M., Krueger, G.: On the Convergence of EM-Like Algorithms for Image Segmentation using Markov Random Fields. Medical Image Analysis 15(6), 830–839 (2011)
11. Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, A., Tappen, M., Rother, C.: A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(6), 1068–1080 (2008)
12. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(11), 1222–1239 (2001)
13. Grady, L.: Multilabel Random Walker Image Segmentation Using Prior Models. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 763–770 (2005)
14. Grady, L.: Random Walks for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(11), 1768–1783 (2006)
15. Marroquin, J., Vemuri, B., Botello, S., Calderon, F., Fernandez-Bouzas, A.: An Accurate and Efficient Bayesian Method for Automatic Segmentation of Brain MRI. IEEE Transactions on Medical Imaging 21(8), 934–945 (2002)
16. Couprie, C., Grady, L., Najman, L., Talbot, H.: Power Watersheds: A Unifying Graph-Based Optimization Framework. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–17 (2010)
17. Li, S.Z.: Markov random field modeling in computer vision. Springer, Berlin (1995)
18. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging 17, 87–97 (1998)
19. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational Methods for Multimodal Image Matching. International Journal of Computer Vision 50(3), 329–343 (2002)
20. Ribes, D., et al.: Comparison of tissue classification models for automatic brain MR segmentation. In: Int. Society for Magnetic Resonance in Medicine (2011)
21. Kwan, R.S., Evans, A., Pike, G.: MRI simulation-based evaluation of image-processing and classification methods. IEEE Transactions on Medical Imaging 18(11), 1085–1097 (1999)
22. Bell, W.N., Olson, L.N., Schroder, J.B.: PyAMG: Algebraic multigrid solvers in Python v2.0, Release 2.0 (2011)

Anatomical Structures Segmentation by Spherical 3D Ray Casting and Gradient Domain Editing

A. Kronman[1], Leo Joskowicz[1], and J. Sosna[2]

[1] School of Eng. and Computer Science, The Hebrew Univ. of Jerusalem, Israel
[2] Dept. of Radiology, Hadassah Hebrew University Medical Center, Jerusalem, Israel
achiak@cs.huji.ac.il

Abstract. Fuzzy boundaries of anatomical structures in medical images make segmentation a challenging task. We present a new segmentation method that addresses the fuzzy boundaries problem. Our method maps the lengths of 3D rays cast from a seed point to the unit sphere, estimates the fuzzy boundaries location by thresholding the gradient magnitude of the rays lengths, and derives the true boundaries by Laplacian interpolation on the sphere. Its advantages are that it does not require a global shape prior or curvature based constraints, that it has an automatic stopping criteria, and that it is robust to anatomical variability, noise, and parameters values settings. Our experimental evaluation on 23 segmentations of kidneys and on 16 segmentations of abdominal aortic aneurysms (AAA) from CT scans yielded an average volume overlap error of 12.6% with respect to the ground-truth. These results are comparable to those of other segmentation methods without their underlying assumptions.

1 Introduction

Patient-specific models of anatomical structures and pathologies generated from volumetric CT/MRI medical images play an increasingly central role in all aspects of patient care, from the initial diagnosis through the planning, delivery, and evaluation of patient treatment. A key task in the generation of these models is the segmentation of anatomical structures and pathologies of interest.

Anatomical structures and pathologies segmentation is a challenging task. Experience shows that each anatomical structure, pathology, and imaging modality has unique characteristics that may lead to significant segmentation errors in a non-negligible number of cases. Correcting faulty segmentations often requires extensive user interaction, trial-and-error parameter tuning, and/or developing custom algorithms, all of which are impractical in a clinical environment.

One of the main reasons that segmentation is challenging is the existence of fuzzy boundaries between the structure of interest and its neighboring structures (Fig. 1a). This occurs when the voxels intensity gradient magnitudes of the target and its neighboring structures boundaries are small, i.e., when their tissue imaging characteristics are similar and when partial volume effect, and/or imaging
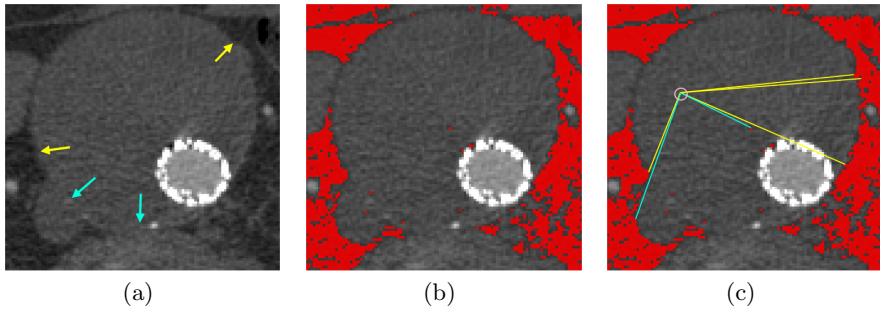
Fig. 1. Illustration on an AAA axial CT slice: (a) original slice – the arrows show sharp (yellow) and fuzzy (cyan) boundaries; (b) result of conservative intensity-based background (red) thresholding to the image; (c) rays cast from a seed point (purple circle) stop at the true boundary (yellow rays) or at a false boundary (cyan rays). Note that for two neighbor rays, their length difference is large when one ray stops at the true boundary and the other ray does not.

noise blurs the image. As a result, the segmentation volume may expand outside the target structure boundary into neighboring structures, thus producing what is termed as segmentation leaks. Fuzzy boundaries cause segmentation leaks in nearly all segmentation methods. In intensity-based thresholding methods, the thresholding classification rule is based on statistically significant intensity differences between neighboring structures, so leaks will appear in fuzzy boundaries. In adaptive region growing [1] and ray casting methods [2], leaks occur when the intensity based stopping criteria does not apply for the fuzzy boundaries voxels. Intensity-based graph-cut methods may produce leaks because the min-cut is unlikely to separate voxels with similar intensities, as this penalizes the global energy function. Energy-based methods, including snakes and level-set active contours, may also produce leaks since they move away the evolving segmentation contour in low-gradient regions because the image term in the energy equation yields a higher overall energy.

Various methods have been proposed to address the fuzzy boundaries problem. The most popular ones incorporate global shape priors [3,4] and curvature-based constraints [5,6] into the segmentation. While global shape priors are useful, they have the following limitations: 1) they can not handle pathologies that have no representative shape; 2) they may not converge for far-from-average cases; and 3) they are difficult to acquire, since they depend on the availability of a delineated atlas. Adding curvature-based constraints into energy-based segmentation methods helps to avoid segmentation leaks when the leak region has a narrow bottleneck shape. However, since the stopping criteria has to be tailored anew for each specific structure, the user often needs to actively stop the segmentation process, which may be inaccurate and inconsistent, and is thus undesirable.

Dodin et al. [2] describe a 2D ray casting method for the segmentation of knee bones in MRI scans. Their rays stopping criteria is based on the Laplacian zero-crossing, so their method will produce segmentation leaks in structures with

extended fuzzy boundaries caused by adjacent structures with similar intensity values. Moreover, 2D ray casting requires at least one seed point for each scan slice, which may be impractical.

In this paper we present a new seeds-based 3D ray casting segmentation method that handles fuzzy boundaries without shape priors or curvature constraints. Its advantages are that: 1) it does not require a global shape prior; 2) it does not assume a bottleneck-like shape of leaks; 3) it requires a few seeds; 4) it does not rely on the user to actively stop the segmentation; 5) it is robust to anatomical variability, noise and segmentation parameters; and, 6) it is applicable to a variety of anatomical structures. Our experimental study on segmentations of 23 kidneys and 16 abdominal aortic aneurysms (AAA) from CT scans yielded an average volume overlap error of 12.6% with respect to their ground-truth segmentations. These results are comparable to other segmentation methods which require shape priors or curvature penalties, but without parameters adjustment and underlying assumptions.

2   Method

The algorithm inputs are one or more seeds located inside the target structure – the seeds are either provided by the user or generated automatically. For each seed point, a segmentation is generated in three steps: 1) initial segmentation by spherical 3D ray casting; 2) identification of correct/incorrect segmentation regions by spherical discontinuation computation; and 3) correction of the faulty boundaries by Laplacian interpolation on the unit sphere. The segmentation results for each seed are combined to produce the final segmentation.

2.1   Initial Segmentation by Spherical 3D Ray Casting

Let $I = \{I_i\}$ be a volumetric image consisting of n voxels I_i. Let $T = \{T_i\}$ be a target structure of interest in I, and let $S = \{S_i\}$ be the unknown surface of T. We divide the image I into two disjoint subsets based on a surface thresholding criteria: the *sharp voxels set* $C = \{C_i\}$, and the *fuzzy voxels set* $F = \{F_i\}$ (Fig. 1b). The thresholding can be based on gray-level values, gradient magnitudes, the output of an edge detector, or any other suitable criteria.

The inputs to the algorithm are k predefined seed points $p_1, ..., p_k$ located inside the target structure (Fig. 2a). For each seed point, the initial segmentation is computed as follows. For each angle $0 \leq \theta_i \leq 2\pi$ and $0 \leq \phi_i \leq \pi$, a 3D ray emanating from the seed in the direction defined by θ_i and ϕ_i is cast until it hits a voxel in C. The resulting segmentation is the total volume covered by the rays. Note that this initial segmentation includes both parts of the true boundary S and also faulty boundary segments resulting from fuzzy boundaries and/or noise (Figs. 2b,3e,3g). We choose this segmentation method because it enables us to automatically estimate and correct the segmentation errors in the next steps.
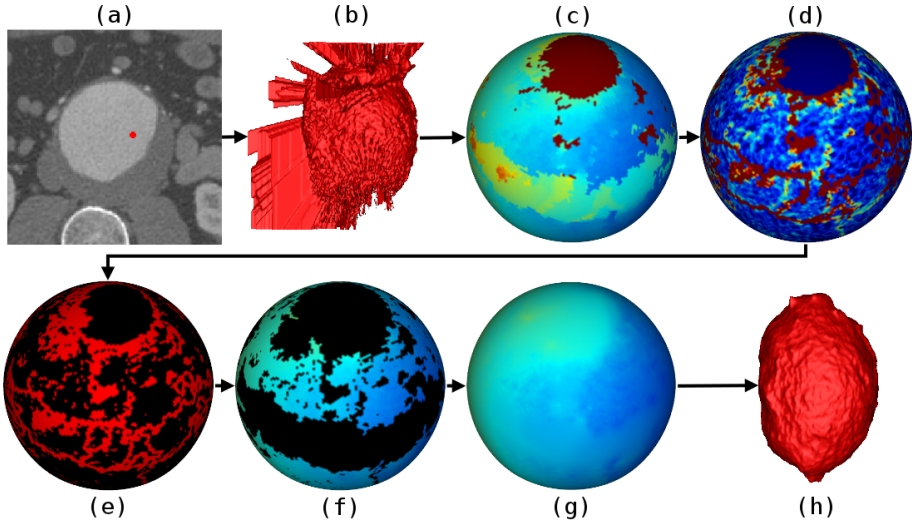
Fig. 2. Illustration of the algorithm: (a) AAA CT axial slice with a seed point (red dot); (b) initial mesh generated by spherical 3D ray casting segmentation; (c) rays length mapping to the sphere – warm (red) colors indicate long rays; (d) rays length gradient magnitude; (e) gradient magnitude thresholding – nodes with large gradient values are red; (f) rays length on the largest connected component; (g) estimation of the unknown boundaries by Laplacian interpolation; (h) final corrected segmentation

2.2   Errors Estimation by Spherical Discontinuation Computation

Next, we classify the cast rays into those that stopped at the true boundary S and those that did not. The estimation is based on the following observation. For each orientation (θ_i, ϕ_i), the rays will stop at the true boundary S with the exception of orientations where there exist a voxel that belongs to C in the interior of T or for directions where the boundary S is fuzzy, i.e. when it does not belong to C. In the first case, the rays will stop before the true boundary, while in the second case the rays will stop after the true boundary (Fig. 1c). As a result, the rays length difference for two voxels in sequential orientations will be relatively large when one of the voxels belongs to the true boundary S and the other does not. In contrast, this difference will be much smaller when both voxels belong to the true boundary.

We formalize this observation to identify the incorrect boundaries of the initial segmentation as follows. Let $f : [0, 2\pi] \times [0, \pi] \to \Re^+$ be a scalar function on the unit sphere such that $f(i, j) = l_{ij}$ where l_{ij} is the length of the ray cast in the direction (θ_i, ϕ_j).

Following [7], we sample f uniformly by constructing a triangulated mesh approximation of the unit sphere by icosahedron subdivision [8] centered around the origin. The result is a geodesic sphere whose nodes are equidistant from each other and have no interpolation singularities around the sphere poles. We associate to

each node its matching f values, $f(Sp_i) = l_{jk}$, where Sp_i is a node in the generated mesh Sp and l_{jk} is the length of the ray cast in the direction defined by Sp_i spherical coordinates (θ_j, ϕ_k) relative to the seed point (Fig. 2c).

Next, we find the sharp changes in f by computing for each mesh node Sp_i its gradient magnitude:

$$|\nabla f(Sp_i)| = \sqrt{\sum_{Sp_j \in N_i} (f(Sp_i) - f(Sp_j))^2} \qquad (1)$$

where N_i are the neighboring nodes of Sp_i (Fig. 2d). Based on the previous observation, we compute the set of nodes $B \subseteq Sp$ that match to the interface between the correct and faulty initial segmentation by thresholding:

$$B = \{Sp_i : \quad |\nabla f(Sp_i)| > t\} \qquad (2)$$

where $t > 0$ is a preset constant with a fixed value for every image (Fig. 2e).

Finally, we isolate the sphere nodes that correspond to orientations with correct initial segmentation by choosing the largest nodes connected component that are separated by nodes in B (Fig. 2f).

2.3   Correct Boundaries Detection by Laplacian Spherical Interpolation

This step detects the true boundaries in the directions where the initial segmentation was incorrect (Fig. 2g). For this we use Laplace interpolation, which minimizes the integrated square of the gradients of f, and thus yields the smoothest surface with respect to this criteria. The Laplace interpolation with Dirichlet boundary condition is formulated as:

$$\nabla^2 f(Sp_i) = 0, \qquad \text{subject to:} \qquad f(Sp_j) = f^*(Sp_j) \qquad (3)$$

for each unknown node Sp_i on the mesh and for each node Sp_j on the boundaries of the interpolation domain with a known value $f^*(Sp_j)$, where $\nabla^2 f(Sp_i) = \sum_{Sp_k \in N_i} (f(Sp_i) - f(Sp_k))$ is the discrete graph Laplacian operator [9], and N_i are the neighbors of node Sp_i on the sphere mesh Sp. The unknown values $f(Sp_i)$ are then computed by solving the resulting sparse system of linear equations.

We generate the final segmentation with the following classification rule to each image voxel I_l at orientations (θ_i, ϕ_j) relative to seed point p_k:

$$L_k(I_l) = \begin{cases} 1 & \text{if } d(p_k, I_l) \leq \hat{f}_{ij} \\ 0 & \text{if } d(p_k, I_l) > \hat{f}_{ij} \end{cases} \qquad (4)$$

where $L_k(I_l)$ is the label for voxel I_l for seed p_k, $d(p_k, I_l)$ is the Euclidean distance between the voxel I_l and the seed point p_k, and $\hat{f}_{ij}$ is the distance interpolation solution for orientations (θ_i, ϕ_j). Figs. 2h,3f,3h show examples of the results.

The key advantage of this new estimation by interpolation method is that any remaining false positive estimation of a leak region caused by choosing to lower

the thresholding parameter t in (2) will be set to its original, correctly estimated value.

Note that the segmentation by ray-casting just described applies to voxels that are visible from the seed point p_k, i.e. to a convex or star-shaped objects. We directly extend the method to general shapes by taking the union of the segmentations L_i obtained for each seed p_i:

$$L_{final}(I_l) = \bigvee_{i=1}^{k} L_i(I_l) \tag{5}$$

The number of seeds required for the structure segmentation depends on the number of its star-shaped components. For the structures of interest, a few seed points are sufficient to segment the structure appropriately.

3   Experimental Results

To quantify the scope, accuracy, and robustness of our method, we designed and conducted the following experimental study. We retrospectively selected 28 CT scans of two different structures, kidneys (organ) and aortic arch aneurysms (AAA – vascular pathology). These structures are different in shape, intensity homogeneity, and surrounding structures. For the kidney study, we included 12 clinical CT datasets (2 kidneys per scan) of size $512 \times 512 \times 350 - 500$ voxels, each of size $0.5 - 1.0 \times 0.5 - 1.0 \times 1.0 - 1.5 \text{mm}^3$, with and without contrast agent, acquired with a Brilliance 64-row CT scanner (Phillips Healthcare, Cleveland, OH). Ground-truth segmentations of the kidneys were obtained from three manual segmentations of experts with STAPLE [10]. For the AAA, we included 16 clinical CT scans of size $512 \times 512 \times 500$ voxels, each of size $0.7 - 1.2 \times 0.7 - 1.2 \times 0.7 - 1.2 \text{mm}^3$ from the same scanner. Imaging streaking artifacts caused by implanted stents were present in several scans (Fig. 3c). An expert radiologist manually generated the ground-truth segmentations.

The method internal parameters values where fixed for all scans to constant values. The background thresholding level was set from the intensity levels, where each voxel with gray-value < 0 HU is a surface candidate. The rays lengths gradient thresholding constant was set to $t = 3$ mm. We found empirically that these values minimize the false negative leaks detection rate. Note that with these parameters setup up to 50% of the target surface was considered as "fuzzy". The number of the icosahedron subdivisions was set to 6. The program runs on a 64-bit quad-core 2.80GHz processors and 6GB memory PC. The seeds were placed interactively by the user: after each seed placement the volume that is seen by the seed was marked on the image until the entire target volume was covered. The mean number of seeds required for the segmentation was 3 (std=0.69) and 1.33 (std=0.5) for the kidney and the AAA. The mean running time was 35.7 (std=2.7) secs and 41.1 (std=10.7) secs for the kidney and the AAA.

To quantify the accuracy of our method, we used both volumetric and surface based metrics. For the kidney, the mean absolute volume difference was 6.02%
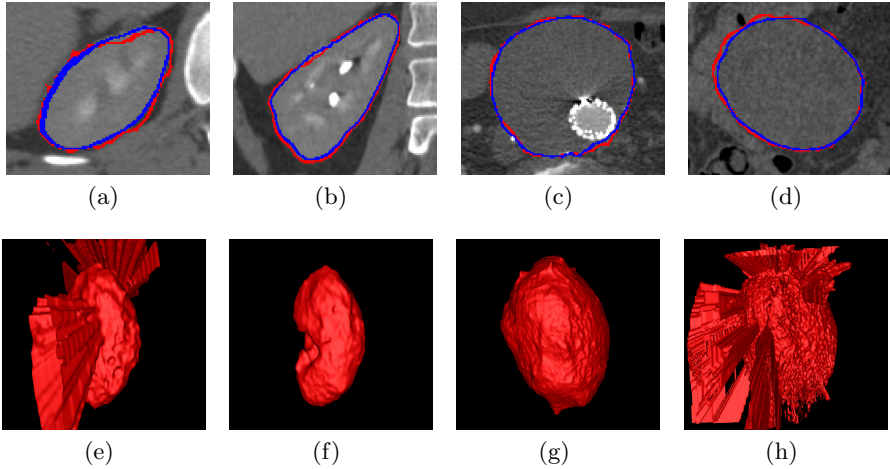
Fig. 3. (a-d): representative results of our segmentation method (red) and manual ground truth (blue): (a) kidney axial CT slice; (b) kidney coronal slice; (c) AAA axial slice; (d) AAA coronal slice; (e-h): 3D meshes of kidney (e-f) and AAA (g-h): (e,g) 3D meshes generated by spherical 3D ray casting; (f,h) final meshes after applying our leaks correction method

(std=4.07%), the mean volumetric overlap error was 12.65% (std=2.46%) and the average symmetric surface distance was 1.12 (std=0.4) mm. For the AAA, the absolute volume difference was 7.99% (std=8%), the mean volumetric overlap error was 12.63% (std=4.35%), and the average symmetric surface distance was 1.29 (std=0.3) mm. Fig. 3 shows a representative example. These results are comparable to previously reported methods [11,12,13,14] without their underlying assumptions and with a significant improvement in the running time.

4 Conclusion

We have developed a new method for the segmentation of anatomical structures in medical images. Our method is based on 3D spherical ray casting and segmentation leaks correction by editing the gradients of the cast rays lengths mapped to the unit sphere. Our experimental study on segmentations of 23 kidneys and 16 abdominal aortic aneurysms from CT scans yielded a mean volume overlap error of 12.6% with respect to the ground-truth. This was deemed clinically acceptable for diagnosis and surgery planning by the co-author expert radiologist.

The advantages of our method are that it does not require a global shape prior, that it does not assume bottleneck-like shape of the leak, that it does not rely on the user to actively stop the segmentation, that it is robust to anatomical variability, noise and segmentation parameters, and that it is applicable to a variety of anatomical structures. Its limitation is that it requires a predefined seed point for each star-shaped component of the target structure. However, for

structures such as the kidney and the AAA, only a few (1-4) seeds are needed in practice. Ongoing and future work includes the evaluation of the method on additional anatomical structures.

References

1. Pohle, R., Toennies, K.: Segmentation of medical images using adaptive region growing. In: Proc. SPIE Medical Imaging, vol. 4322, pp. 1337–1346 (2001)
2. Dodin, P., Martel-Pelletier, J., Pelletier, J., Abram, F.: A fully automated human knee 3d mri bone segmentation using the ray casting technique. In: Medical and Biological Engineering and Computing, pp. 1–12 (2011)
3. Freedman, D., Zhang, T.: Interactive graph cut based segmentation with shape priors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 755–762 (2005)
4. Tsai, A., Yezzi Jr., A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. IEEE Transactions on Medical Imaging 22(2), 137–154 (2003)
5. Djabelkhir, F., Khamadja, M., Odet, C.: Level set constrained segmentation using local curvature. In: 5th International Symposium on Image and Signal Processing and Analysis, ISPA, 152–155. IEEE (2007)
6. El-Zehiry, N., Grady, L.: Fast global optimization of curvature. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3257–3264 (2010)
7. Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J.J., Shenton, M.E., Gerig, G.: Framework for the statistical shape analysis of brain structures using spharm-pdm. Insight J. (1071), 242–250 (2006)
8. Loop, C.: Smooth subdivision surfaces based on triangles (1987)
9. Reuter, M., Biasotti, S., Giorgi, D., Patanč, G., Spagnuolo, M.: Discrete laplace-beltrami operators for shape analysis and segmentation. Computers & Graphics 33(3), 381–390 (2009)
10. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging 23(7), 903–921 (2004)
11. Freiman, M., Kronman, A., Esses, S., Joskowicz, L., Sosna, J.: Non-parametric Iterative Model Constraint Graph min-cut for Automatic Kidney Segmentation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 73–80. Springer, Heidelberg (2010)
12. Rao, M., Stough, J., Chi, Y., Muller, K., Tracton, G., Pizer, S., Chaney, E.: Comparison of human and automatic segmentations of kidneys from ct images. Int. J. Radiat. Oncol. Biol. Phys. 61(3), 954–960 (2005)
13. Lin, D., Lei, C., Hung, S.: Computer-aided kidney segmentation on abdominal ct images. IEEE Transactions on Information Technology in Biomedicine 10(1), 59–65 (2006)
14. Freiman, M., Esses, S.J., Joskowicz, L., Sosna, J.: An iterative model-constraint graph-cut algorithm for abdominal aortic aneurysm thrombus segmentation. In: Proc. of the 2010 IEEE Int. Symp. on Biomedical Imaging, ISBI 2010 (2010)

Segmentation of the Pectoral Muscle in Breast MRI Using Atlas-Based Approaches

Albert Gubern-Mérida[1], Michiel Kallenberg[2],
Robert Martí[1], and Nico Karssemeijer[2]

[1] University of Girona, Spain
{agubern,marly}@eia.udg.edu
[2] Radboud University Nijmegen Medical Centre, The Netherlands
{m.kallenberg,n.karssemeijer}@rad.umcn.nl

Abstract. Pectoral muscle segmentation is an important step in automatic breast image analysis methods and crucial for multi-modal image registration. In breast MRI, accurate delineation of the pectoral is important for volumetric breast density estimation and for pharmacokinetic analysis of dynamic contrast enhancement. In this paper we propose and study the performance of atlas-based segmentation methods evaluating two fully automatic breast MRI dedicated strategies on a set of 27 manually segmented MR volumes. One uses a probabilistic model and the other is a multi-atlas registration based approach. The multi-atlas approach performed slightly better, with an average Dice coefficient (DSC) of 0.74, while with the much faster probabilistic method a DSC of 0.72 was obtained.

Keywords: pectoral muscle, breast MRI, atlas-based segmentation.

1 Introduction

Automatic identification of pectoral muscle is an important step in methods for automatic breast cancer assessment in most image modalities. For instance, in mammography, the most used image modality in screening programs, the detection and removal of the pectoral muscle is often used to remove false positive marks of Computer Aided Detection (CAD) systems [3]. In Magnetic Resonance Imaging (MRI) of the breast, the image modality employed in the presented work, the importance of the pectoral muscle detection has recently been recognized in two applications. Firstly, breast tissue density has been identified as an important risk factor for developing breast cancer, being four times larger in women with a breast density higher than 75%, compared to those with little or no density [2]. Breast MRI provides a good tissue contrast between fibroglandular and fatty tissues and a three-dimensional characterization of breast composition. These good properties in the breast tissue have been a strong reason to use breast MRI in breast density measurement [9,8,5]. However, the contrast between pectoral muscle and dense tissue is poor. Hence, a first step to separate the breast from the body is commonly essential. This separation is not trivial

due to the large shape and intensity variations in the pectoral muscle of different patients. Some solutions are present in the literature: boundary tracing or spline fitting without [10] and with manual intervention [9,8] and delineation of the whole breast using breast models [4], but none of them completely delineates the pectoral muscle.

Secondly, MRI is often used with a contrast agent for lesion detection. For a better interpretation of contrast enhancement lesions, researchers have tried to incorporate pharmacokinetic modeling to the interpretation of the MRI. Some of these models require calibrations with respect to reference tissues and make use of the signal intensity of specific regions for determining physiological measures [11]. In breast MRI, the pectoral muscle can be used as a reference tissue given its properties.

Atlas-based segmentation has been shown to be a powerful technique for automatic delineation of anatomical structures in different 3D image modalities [1,6]. Multi-atlas and probabilistic approaches are the most commonly used strategies. By definition, the former is supposed to obtain more precise segmentations than the latter. However, multi-atlas approach is far more time consuming. There has been only one initial attempt that uses an atlas strategy for breast MRI segmentation [5], but the segmentation of the pectoral muscle was not the main interest of the work. Moreover, the method followed a probabilistic approach using one reference atlas, which could have some limitations. As shapes are highly variable, the reference choice affects final results.

The novelty of this paper consist in the study of fully automatic atlas-based methods for pectoral muscle segmentation in breast MRI in terms of performance and complexity. A dedicated multi-atlas approach based on [6] is proposed (see section 3.3) and compared to the probabilistic approach of [5] (see section 3.2). An original breast MRI registration framework focused on the body has been also defined and used in both methods (see section 3.1). To our knowledge, no similar studies are found in the literature. Advantages and inconveniences of both strategies are discussed in sections 4 and 5 and a solution to obtain a reasonable time-accuracy trade-off is proposed.

2   Material

The data set used to evaluate the segmentation results and build the atlases consists of 27 pre-contrast T1-weighted MR breast scans obtained from different patients. Breast MRI examinations were performed on a 1.5 T system (Siemens 1.5T, Magnetom Vision), with a dedicated breast coil (CP Breast Array, Siemens, Erlangen). The pixel spacing differed between volumes with values ranging from 0.625 mm to 0.722 mm. The slice thickness was 1.3 mm and the volume size was 512 x 256 x 120 voxels. Patients were scanned in prone position.

Three experienced observers performed manual segmentations. Two of them manually segmented only the pectoral muscles of 8 cases. The third one manually segmented each of the 27 MR volumes into 7 classes: background, fatty tissue, glandular tissue, pectoral muscles, lung area and the heart. The seventh class is the "other" class and refers the previous non-labeled voxels of the thorax.

Annotations were done every 5-10 slices and linear interpolation was applied to obtain the complete labeling. When needed, and especially for heart, lungs and pectoral muscles, accurate manual delineation was performed with a smaller slice interval. For the manual segmentation of background, fatty and fibroglandular tissue, thresholding was applied over regions of interest provided by the reader. Fig. 1 shows an example of a MRI slice on an axial view and the manual delineation of the mentioned classes. One should note the complexity of performing such ground truth annotations, where each volume takes approximately 45 minutes in a dedicated breast MRI annotation environment.
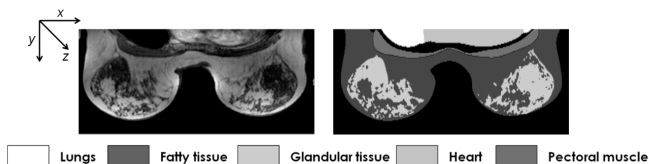
Fig. 1. Breast MR scan on an axial slice with the manual annotation of the different structures

3   Methods

Atlas-based strategies are characterized by the use of prelabeled images, usually manually obtained, to perform the automatic segmentation of new images, also called targets. They employ registration algorithms, which play an important role for the final segmentation. Section 3.1 explains the mapping algorithm used by the two atlas-based approaches evaluated in this paper. Section 3.2 briefly describes the construction and the use of a probabilistic atlas in a Bayesian framework segmentation [5]. Finally, in section 3.3 we report the proposed multi-atlas segmentation algorithm based on [6] for the delineation of the pectoral muscle in breast MRI.

3.1   Registration

Registration is an important step in atlas-based segmentation algorithms. Without an accurate transformation between the structures we aim to segment, the segmentation can not perform accurately. For this reason we developed a registration framework focused on the body area. We observed that the sternum is always localized between pectoral muscles. Hence, by accurately localizing the sternum the pectoral muscles can be aligned. Our registration approach is initialized by detecting the sternum of the subjects. Automatic sternum detection is based on [5]. Then, the volumes are cropped at 2 cm distance anterior to the sternum position to focus the registration on the area of the body. By doing so, most breast tissue is removed and can not negatively bias the final mapping of body structures. The 2 cm distance anterior to the sternum ensures that pectoral muscle voxels are not discarded.

The registration process is composed by two stages. First, a translation transform is performed, where translation along the y axis is defined by the distance between y-coordinates of both sternums. Translation along x and z axis is found by optimizing the similarity measure. The second stage is a non-rigid transform based on B-Splines registration in a multi-resolution scheme using a stochastic gradient descent optimizer. Three resolutions were defined. B-Splines grid spacing was set to 32, 16 and 8 mm for each of the resolutions taking the size of the pectoral muscle into account. The similarity measure maximized by the whole framework was normalized cross correlation (NCC) as all the datasets were acquired with the same modality. Elastix [7] was used for the implementation.

3.2   Method 1: Probabilistic Atlas-Based Segmentation

In the presented atlas-based segmentation method, a probabilistic atlas is used in a Bayesian framework to provide an accurate probability distribution for the pectoral an the thoracic area. Following a leave-one-out evaluation strategy, for each patient segmentation, a full probabilistic atlas was built offline with the 26 remaining patients. These 26 patients and their segmentations were mapped using the registration method explained previously into the same reference space. The probabilistic atlas was created by computing the frequency with which each location was labeled as a specific organ. A common reference space was used for all the experiments by visually selecting an extra patient which has normal appearance. The reference case, or anatomical image of the atlas, was not included in the evaluation set.

Figure 2 shows the general schema of the segmentation framework with Bayesian voxel classification algorithm incorporating the use of the probabilistic atlas. From the top to bottom, the probabilities of the atlas are mapped by registration of section 3.1 onto target image space {T} using the anatomical image
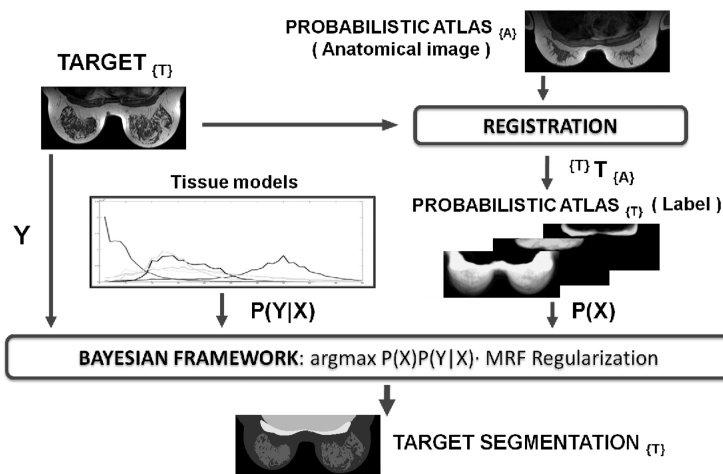
Fig. 2. Probabilistic atlas segmentation approach overview

of the atlas. The probabilistic atlas, the tissue models (previously built from the scans and manual segmentations of the data set) and the target are provided to the Bayesian framework as a prior probability $P(X)$, conditional probability $P(Y|X)$ and set of intensity values Y, respectively. The Bayesian framework estimates the segmentation X that maximizes $P(X)P(Y|X)$ and also includes a Markov Random Field (MRF) regularization to smooth the segmentation taking neighborhood information into account [5].

3.3   Method 2: Multi-atlas Segmentation

Multi-atlas segmentation approaches consider all the volumes of the dataset and their manual segmentations as individual atlases. The term atlas is defined as the pair of the anatomical image (MRI volume) and its manual segmentation or label. The process to obtain an automatic segmentation for a target volume is illustrated in figure 3. First, given the target volume T, all the atlases are mapped onto the target space using the registration algorithm of section 3.1. Subsequently, the deformed anatomical images are compared to the target to perform a selection of the most similar atlases. The selection is based on the Normalized Cross Correlation similarity measure and a ratio defined as follows:

$$r_i = \frac{NCC(T, A_i \circ M_i)}{max_j \, NCC(T, A_j \circ M_j)}, \tag{1}$$

where M refers to the mapping between the target and an atlas and j refers to the deformed atlas with maximum similarity. An atlas A_i is selected if it satisfies $r_i \geq \varphi$. A value of $\varphi = 0.9$ empirically appeared to be the best value for our results.
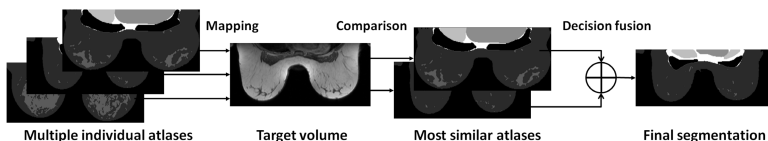
Fig. 3. Multi-atlas segmentation approach overview

Finally, the selected deformed atlas labels are fused to yield a single final segmentation of the patient or target image. This step is called decision fusion and defines how the deformed segmentations of the selected atlas are combined. In this work we have made use of majority voting method, which was proven to give good results in [6].

4   Results

In a leave-one-out experiment we evaluated the probabilistic and the multi-atlas segmentation frameworks on 27 patients. Each segmented case was not included

for the construction of the probabilistic atlas or within the set of individual atlases respectively. The quality of the segmentation was measured by determining the similarity of the segmentation with the ground truth using the Dice Similarity Coefficient (DSC). DSC was chosen as it is commonly used in the literature [6,5]. For all cases we manually discarded initial and last slices which do no contain relevant information or are clearly affected by noise. Figure 4(a) shows a box plot with DSC values for each method. Segmentation results are similar (DSC median of 0.76 for both and DSC mean $\pm$ sd of 0.72 ± 0.09 and 0.74 ± 0.06 for probabilistic and multi-atlas respectively), but multi-atlas framework slightly outperforms the probabilistic. These results can be better seen in figure 4(b), where DSC values of each case using both methods are shown.

(a)   (b)
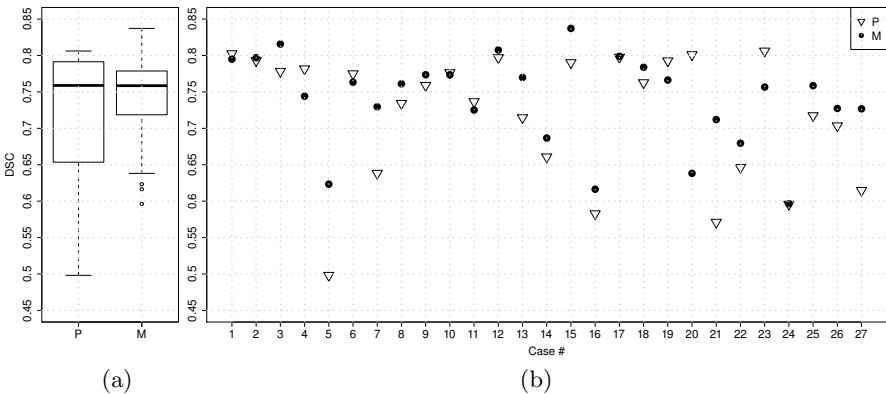
Fig. 4. (a) Box plot with segmentation DSC for pectoral segmentation and (b) DSC segmentation results for each of the 27 cases using probabilistic (P) and multi-atlas (M) approaches

Lower DSC values are mainly due to the registration process not being able to compensate the differences between volumes. This is more the case of the probabilistic approach as the method uses only one registration with a single reference. In those cases (see case 7 for instance), the multi-atlas approach performs better as it includes multiple registrations and selects the best ones. Only in one case (number 20) the probabilistic approach obtains results much better than multi-atlas, where pectoral muscle segmentation in initial slices is not really precise (labeled as thorax instead). However, accurate delineations in intermediate slices are obtained for both methods as it is illustrated in figure 5, where three examples of automatic and manual segmentations are shown.

Finally, since no previous works performed pectoral segmentation in breast MRI, inter-observer variability was computed by 3 viewers over 8 manual segmentations. DSC mean of 0.70 ± 0.12 and median of 0.72 were obtained, lower than the DSC values achieved by the automatic atlas-based approaches.

All the tests have been launched on Intel(R) Core(TM)2 Quad CPU Q9550 2.83GHz. Starting with the common step, registration between two volumes takes

Multi-atlas Probabilistic Manual
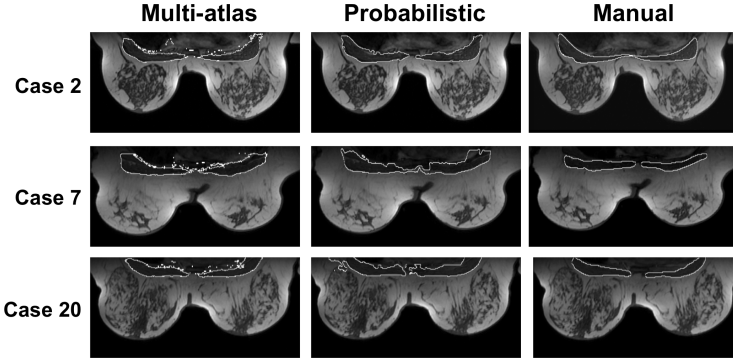
Fig. 5. Intermediate slices from 3 different cases and their segmentations

$t_r \approx 12$ min. The complexity time for multi-atlas segmentation is explained as $t_{multi-atlas} \approx N \times (t_r + t_a) + n \times t_f$, where N is the number of individual atlases ($N = 26$), t_a the time to compute the mapping and the comparison of an individual anatomical atlas ($t_a \approx 4$ min.), n the number of selected atlases and t_f the time to propagate and fuse an atlas labeled image ($t_f \approx 3$ min.). In the best scenario, being only one atlas selected ($n = 1$), the computation time to obtain a segmentation using a multi-atlas approach is $t_{multi-atlas} \approx 419$ minutes (7 hours). The complexity time for the probabilistic approach is defined as $t_{probabilistic} \approx t_r + t_p + t_b$, where t_p is the time to map the probability distributions to the target space ($t_p \approx 8$ min) and t_b the time to perform the segmentation based on Bayesian theory ($t_b \approx 10$ min). Approximately, $t_{probabilistic} \approx 30$ min.

5   Discussion

In this work, the atlas-based methodology has been studied to perform the complete delineation of the pectoral muscle in breast MRI, which has not been done previously. Fully automatic and dedicated multi-atlas and probabilistic frameworks have been proposed and tested on 27 different patients.

The obtained results are satisfactory in both frameworks, with DSC values higher than the computed inter-observer variability. It proves the high reliability of atlas-based segmentation methods to perform pectoral delineations. However, we are aware that the evaluation and the construction of the atlases were performed with annotations from a single viewer, as obtaining 3-dimensional manual segmentations is a time consuming task. The low inter-observer DSC value explains the difficulty and subjectivity to delineate the pectoral muscle. Its shape has high-variability and cartilage, intercostal muscles and fatty tissue also appear in the area. The inclusion of these tissues depends on the observer opinion.

As it was expected, multi-atlas segmentation appears to be more consistent than probabilistic. This is explained by the fact that multi-atlas approach includes an atlas selection step to choose the most similar atlas compared to the

segmented volume. In the probabilistic framework, when the target differs considerably from the reference atlas and the registration can not compensate the differences, the final segmentation becomes affected with slightly poorer results. However, the computation time for a multi-atlas segmentation is 14 times larger.

Considering the influence of atlas selection, in future work we will study a multi-probabilistic atlas framework. A larger dataset will be created with annotations from different observers. We will group different breast MRI populations based on shape. For each population, a probabilistic atlas will be built. The most similar atlas to the image at hand will be chosen for segmentation.

References

1. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J., Rueckert, D.: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. NeuroImage 46(3), 726–738 (2009)
2. Boyd, N., Guo, H., Martin, L., Sun, L., Stone, J., Fishell, E., Jong, A., Hislop, G., Chiarelli, A., Minkin, S., Yaffe, M.: Mammographic density and the risk and detection of breast cancer. N. Engl. J. Med. 356, 227–236 (2007)
3. Camilus, K.S., Govindan, V.K., Sathidevi, P.S.: Computer-aided identification of the pectoral muscle in digitized mammograms. Journal of Digital Imaging 23, 562–580 (2010)
4. Gallego, C., Martel, A.L.: Automatic model-based 3d segmentation of the breast in mri. In: Proc. of SPIE 2011, vol. 7962 (2011)
5. Gubern-Mérida, A., Kallenberg, M., Martí, R., Karssemeijer, N.: Fully automatic fibroglandular tissue segmentation in breast mri: an atlas-based approach. In: MICCAI 2011: Workshop on Breast Image Analysis pp. 73–80 (2011)
6. Klein, S., van der Heide, U., Lips, I., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information. Medical Physics 35(4), 1407–1417 (2008)
7. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: Elastix: a toolbox for intensity based medical image registration. IEEE Transactions on Medical Imaging 29(1), 196–205 (2010)
8. Klifa, C., Carballido-Gamio, J., Wilmes, L., Laprie, A., Shepherd, J., Gibbs, J., Fan, B., Noworolski, S., Hylton, N.: Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. Magnetic Resonance Imaging 28(1), 8–15 (2009)
9. Nie, K., Chang, D., Chen, J., Shih, T., Hsu, C., Nalcioglu, O., Su, M.: Impact of skin removal on quantitative measurement of breast density using mri. Med. Phys. 37(1), 227–233 (2010)
10. Wang, L., Filippatos, K., Friman, O., Hahn, H.K.: Fully automated segmentation of the pectoralis muscle boundary in breast mr images. In: Proc. of SPIE 2011, vol. 7963 (2011)
11. Yankeelov, T.E., Luci, J.J., Lepage, M., Li, R., Debusk, L., Lin, P.C., Price, R.R., Gore, J.C.: Quantitative pharmacokinetic analysis of dce-mri data without an arterial input function: a reference region model. Magn. Reson. Imaging 23(4), 519–529 (2005)

Hierarchical Conditional Random Fields for Detection of Gad-Enhancing Lesions in Multiple Sclerosis

Zahra Karimaghaloo[1], Douglas L. Arnold[3],
D. Louis Collins[2], and Tal Arbel[1]

[1] Centre for Intelligent Machines, McGill University, Canada
[2] Montreal Neurological Institute, McGill University, Canada
[3] NeuroRx Research, Montreal, Canada

Abstract. The detection of gad-enhancing lesions in brain MRI of Multiple Sclerosis (MS) patients is of great interest since they are important markers of disease activity. However, many of the enhancing voxels are associated with normal structures (i.e. blood vessels) or noise in the MRI, making the detection of gad-enhancing lesions a challenging task. Furthermore, these lesions are typically small and in close proximity to vessels. In this paper, we present an automatic, probabilistic Hierarchical Conditional Random Field (HCRF) framework for detection of gad-enhancing lesions in brain images of patients with MS. In the first level, a CRF with unary and pairwise potentials is used to identify candidate lesion voxel. In the second level, these lesion candidates are grouped based on anatomical and spatial features, and feature-specific lesion based CRF models are designed for each group. This lesion level CRF incorporates higher order potentials which account for shape, group intensities and symmetries. The proposed algorithm is trained on 92 multimodal clinical datasets acquired from Relapsing-Remitting MS patients during multi-center clinical trials and is evaluated on 30 independent cases. The experimental results show a sensitivity of 98%, a positive predictive value of 66% and an average false positive count of 1.55, outperforming the CRF and MRF frameworks proposed in [1].

1 Introduction

Multiple Sclerosis (MS) is a disorder of the central nervous system which is characterized by focal, inflammatory lesions of the white matter (WM) which appear hyperintense on T2-weighted MRI (i.e. *T2w lesions.*) A subset of MS lesions showing active inflammation can be identified in T1-weighted (T1w) scans acquired after injection of gadolinium-based contrast agents (i.e. *gad-enhancing lesions*). The number and volume of gad-enhancing lesions are important biomarkers of disease activity and can be used in the development of drugs for MS. Gad-enhancing lesions are generally segmented manually, a laborious task subject to intra- and inter-rater variability. As clinical trials for MS treatments usually involve enormous amounts of data from multiple centers, it is desirable

(a) T1c   (b) Binary mask (c) Manual labels (d) Active lesion (e) Manual labels
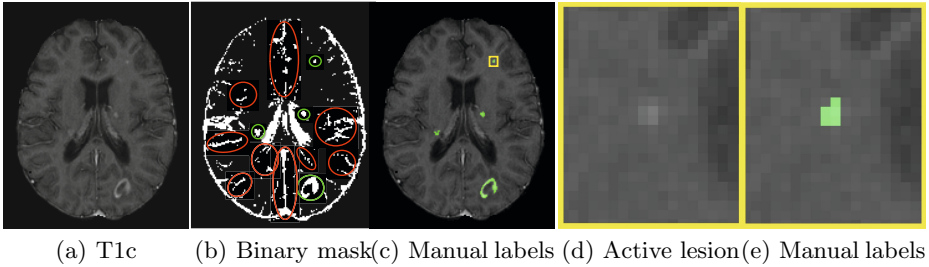
Fig. 1. (a), (b) and (c) respectively show the post-contrast T1w (T1c), the binary mask of the voxels having sufficient enhancement to be considered as lesions and the manual labels (in green). In (b), red and green color mark some non-lesional and lesional enhancements respectively. (d) and (e) show zoomed in images of one of the active lesions without and with the label, respectively.

to have an automatic segmentation method that is robust to data variability due to different scanners and protocols. Unfortunately, automatic identification of enhancing lesions is very challenging. This is due, in part, to variability in size (as small as 3 voxels), texture, intensity and location across patients. Furthermore, the majority of the enhancing voxels are associated with non-lesional, normal structures (e.g. mostly blood vessels) and MRI noise (Fig.1). These factors make the clinical objective of detecting *all* of the gad-enhancing lesions while maintaining low False Positive (FP) counts very difficult. Existing methods for gad-enhancing lesion segmentation are either not fully automatic [2,3], depend on non-standard MRI acquisition sequences [3,4], or require the prior segmentation of T2w lesions in order to remove FPs [4,5]. A Conditional Random Fields (CRF) [1] classifier was recently developed for this task, without relying on the pre-segmentation of T2w lesions and using only commonly acquired MRI sequences (i.e. pre- and post-contrast T1w, T2w, PDw and FLAIR). The CRF was shown to outperform standard MRF, SVM and linear regression models in terms of FPs, virtually eliminating the number of False Negative (FN) lesions. However, although some shape information was used, this model used mainly local, voxel-level unary and pairwise potentials for classification, and was not powerful enough to remove all the FP lesions. Higher order potentials can express more complex image features [6]. However, incorporating them is computationally expensive when considering all enhancing voxels.

In this work, we introduce a probabilistic Hierarchical CRF (HCRF) model for the automatic detection of gad-enhancing lesions which allows context to be incorporated at multiple levels sequentially. In particular, our framework includes: (1) *a voxel level CRF* with unary and pairwise potentials to obtain lesion candidates through a voxel-based classification, and (2) *a regional, lesion level CRF* incorporating higher order potentials, allowing for the integration of shape based features (e.g. elongation indicating vessels), groupwise intensity characteristics and global features (e.g. symmetry).

The experimental results of applying the HCRF classifier on real, multi-centre clinical trial images acquired from 122 patients with Relapsing Remitting MS (RRMS) yields a 98% sensitivity rate, 0.66 positive predictive value (PPV) and an average of 1.55 FP counts per patient when compared to a set of "silver standard" manual labels attained by expert consensus[1]. The results also show that the new HCRF method outperforms standard MRF and previous CRF methods.

2 Method

The goal of image classification is to assign to each voxel in the image a label. Let us denote $\mathbf{x}_i \in \Re^d$ as the observation vector at the voxel i and $y_i \in \{1, 0\}$ as the label of the ith instance. We now describe the 2 main levels of inference in our framework.

2.1 Voxel Level CRF

At the first level of inference, we develop a Conditional Random Field (CRF) [7] classifier at the voxel level. A CRF is a discriminant graphical model that directly estimates the parameters of the conditional posterior, $p(Y|X)$, by learning a mapping from observations to class labels, generally formulated as:

$$p(Y|X) = \frac{1}{Z} \exp(\sum_{i=1}^{n} \phi_v(y_i|X) + \sum_{i,j,i\neq j} \varphi_v(y_i, y_j|X)) \qquad (1)$$

where $X = \{\mathbf{x}_i\}_1^n$ and $Y = \{y_i\}_1^n$. n and Z indicate the total number of pixels and the normalization term, respectively. ϕ_v and φ_v are the *unary* and the *pairwise* potentials at the voxel level, evaluating the likelihood of a voxel taking a particular label and a pair of neighbouring voxels taking on different labels, respectively. The unary potential is modeled as:

$$\phi_v(y_i|X) = \log\ p(y_i|\mathbf{x}_i) = \log(\sigma(g(\mathbf{x}_i))) \qquad (2)$$

where $\mathbf{x}_i$ is the observation vector at voxel i (e.g. intensity values from different MRI sequences) and σ denotes the sigmoid function. Similar to [1], we use a Relevance Vector Machine (RVM) classifier to model $p(y_i|\mathbf{x}_i)$.

The pairwise potential in a CRF model permits learning the relationships between the labels of neighbouring nodes given the observed data. This potential is usually modeled based on the absolute difference in neighbouring observations. In this work, we incorporate the sign of the difference as well in order to account for the directional relationship between the enhancing lesion voxels and the surrounding non-enhanced voxels (e.g. lesion voxels should typically be hyper intense in T2w and hypo intense in T1w pre-contrast comparing to the surrounding non-enhanced voxels). Therefore, to evaluate the likelihood of an

---

[1] It is important to note that errors in manual labeling are very likely present.

observed edge, not only should the gradient in the associated observations be high but the sign of the gradient should also support the labelling configuration. We model the pairwise potential as:

$$\varphi_v(y_i, y_j | \mathbf{x}_i, \mathbf{x}_j) = -[\alpha_{y_i y_j} f_{y_i y_j}(\mathbf{x}_i, \mathbf{x}_j) + \beta_{y_i y_j}](1 - \delta(y_i - y_j)) \qquad (3)$$

where $\alpha_{y_i y_j}$ and $\beta_{y_i y_j}$ are model parameters and δ is the Kronecker delta function. The function $f_{y_i y_j}$ is the data dependant term.

Training at the voxel level of inference is adapted to eliminate the possibility of having any FNs, at the expense of additional FPs. A lesion level CRF will now be developed to remove the remaining FPs based on higher level spatial and structural information.

2.2   The Regional Lesion-Based CRF

After the first level of CRF, voxels with the same label are grouped together to form lesion candidates (each lesion candidate is surrounded by a bounding box, B). As lesion characteristics are highly non-homogeneous throughout the image volume, lesion candidates are grouped in accordance with their spatial and anatomical characteristics. Features are specifically chosen in order to perform this grouping. Once grouped, a feature specific CRF model with the incorporation of higher order potentials is learned for each grouping to optimally distinguish false positive lesion candidates from true lesions.

The spatial and anatomical features used for the grouping includes: average location of each lesion candidate along the axis perpendicular to the axial plane (z-axis), average WM and partial volume (PV) values of each region obtained from spatial probabilistic atlases registered to each patient. The range of these three features are divided to three parts yielding 27 groups overall. Empirically, it was observed that these features are able to efficiently represent the diversity of lesion candidates. After grouping lesion candidates based on these features, we then learn a feature specific CRF model for each group as follows:

$$p(Y|X) = \frac{1}{Z} \exp\left(\sum_{i=1}^{n} \phi_r(y_i|X) + \sum_{i,j,i \neq j} \varphi_r(y_i, y_j|X) + \sum_{b \in B} \psi_r(y_b|X)\right) \qquad (4)$$

where ϕ_r and φ_r are modeled similar to Eq. 2 and 3, respectively. The higher order potential takes the form of the robust P^N model (for details on model, see [6]):

$$\psi_r(y_b|X) = \begin{cases} N(y_b)\frac{1}{Q}\gamma_{max} & \text{if } N(y_b) \leq Q \ . \\ \gamma_{max} & \text{otherwise.} \end{cases} \qquad (5)$$

where b denotes the voxels within the bounding box, B, and $N(y_b)$ denotes the number of voxels in B not taking on the dominant label. Q is the truncation parameter which controls the rigidity of the higher order potential. γ_{max} is determined by evaluating the quality of the lesion candidate within each bounding box and is based on the average intensity values of the region, the shape of the

underlying enhanced structure (to distinguish enhancing lesions from vessels) and assessing if the enhanced region has any symmetrical pair (to distinguish FPs causing from the enhancements at the choroid plexus regions). For lesion candidates where the assessed quality of the candidate is high, γ_{max} has a lower value allowing for the voxels within the bounding box to take on different labels. Otherwise, γ_{max} is high and the model favours all voxels within the bounding box to take on the same label (i.e. we favoured label zero to eliminate FPs).

We use the piecewise learning approach, to learn the parameters of the model [8]. This technique makes the computation tractable by breaking the model into disjoint pieces, each of which is trained independently and integrated afterwards. The parameters of the unary potential at both levels are learned within a Bayesian inference, while the parameters of the pairwise model are learned within a cross validation method. The parameters of the higher order potential are manually selected to minimize the error on the training set. In the inference stage, considering the CRF model at each level and its learned parameters, we seek the most probable labelling that maximizes the conditional probability of Eq. 1 and 4. Graph Cuts are chosen to solve this optimization problem primarily because of their ability to find globally optimal solutions for binary classifications [9][2].

3   Experiments and Results

3.1   Data Pre-processing

The data was acquired from 122 patients with RRMS as part of a multi center clinical trial (31 centers). The patients had varying levels of gad-enhancing lesion loads, located in different areas of the brain WM, and showed varying amounts of brain atrophy. All MRI volumes consist of 3mm thick axial slices with 1mm $\times$ 1mm intra-plane resolution. Each acquisition was composed of five sequences: pre- and post-contrast T1w, T2w, PDw and FLAIR. For the particular data set that we had access to, the "silver standard" manual labels were determined using a protocol where two trained experts label the data separately, followed by consensus agreement. Prior to classification, pre-processing steps including bias-field inhomogeneity correction as well as removal of non-brain regions from the MRI are performed. Furthermore, all intra-subject sequences are registered to a common coordinate space and the intensity histogram of all sequences is normalized [10].

The HCRF classifier is trained on 92 randomly selected MRI volumes and tested on the remaining 30 cases. There is at least one patient from each center in the training set[3]. Fig. 2 shows an example of the performance of the various components of the HCRF framework. While the classification results obtained from the unary potential and unary + pairwise potentials at the voxel level show many FPs, the proposed HCRF model in (2(e)) successfully captured all four enhancing lesions without any FPs.

---

[2] We used the Matlab wrapper for robust higher order potentials by Shai Bagon: http://www.wisdom.weizmann.ac.il/$\sim$bagon/matlab.html.

[3] We had insufficient data from each center to provide a statistical analysis per center.

| (a) T1c | (b) Binary mask | (c) ϕ_v | (d) $\phi_v + \varphi_v$ | (e) HCRF |
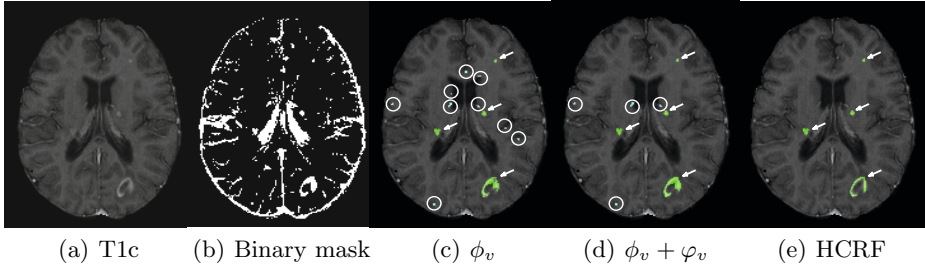
Fig. 2. The performance of the different components of the proposed HCRF classifier. The shown images are: the post contrast T1w (a), the binary mask (b), the classification results of using only the unary potential, $\phi_v(y_i|X)$ (c), using both the unary and pairwise potentials, $\phi_v(y_i|X) + \varphi_v(y_i, y_j|X)$ (d), and the HCRF model (e). Green and cyan show TP and FP regions respectively.

We compare the performance of the HCRF model with the MRF and CRF models described in [1]. Here, the MRF is modeled similar to the CRF model at the voxel level where the pairwise potential is modeled as: $\varphi_v^{MRF}(y_i, y_j) = \beta y_i y_j$ and β is the Ising parameter. In [1], a post-processing step is used to remove any regions with fewer than 3 connected voxels. We found that this can result in missing small lesions, specifically when only 1 or 2 voxels of the lesion are captured by the model. Therefore, there is no post-processing step in our HCRF model. Instead, we try to effectively learn the neighbourhood characteristics of small lesions through the higher order potentials to better distinguish between lesional enhancements and FPs. Fig. 3 shows a qualitative example of the results for a case with one small enhancing lesion in the shown slice, which only the HCRF framework successfully captures. It is missed by the MRF due to the over-smoothing effect of the Ising model, and because it does not consider the interaction of neighbouring observations. For CRF [1], even though 2 voxels of the lesion were captured, the aforementioned post-processing step has removed the lesion.

The performance of our HCRF model is also quantitatively evaluated against the MRF and CRF [1] in Table 1. For a fair comparison, the CRF [1] without the aforementioned post-processing step ($CRF[1]_{no\ post}$) is also included. Comparisons are based on: the sensitivity rate ($\frac{TP}{TP+FN}$ for each scan), the average FP number and the average PPV ($\frac{TP}{TP+FP}$ for each scan). Note that the goal is high lesion detection rates and not the exact delineation of lesion boundaries. Hence, if only one voxel of an enhancing lesion is captured, that is counted as a TP[4]. If all of the voxels of an enhancing lesion are missed, that is counted as an FN and any candidate that does not correspond to an enhancing lesion is counted as an FP. The results show that the HCRF model has the lowest FP count and the highest sensitivity rate over all methods. The pairwise potential in the first

---

[4] For the detected lesions, 70% of the area of enhancing lesions were captured on average.

(a) T1c   (b) Binary mask   (c) MRF   (d) CRF [1]   (e) HCRF
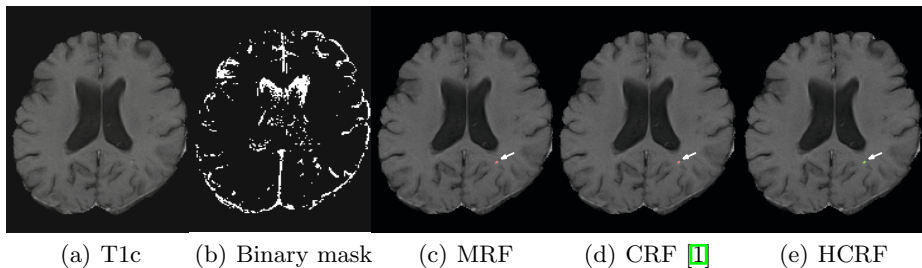
Fig. 3. Comparison of the performance of the HCRF classifier against the MRF and CRF [1] models. The shown images are: the post-contrast T1w (a), the binary mask (b), the classification results of the MRF (c), CRF (d) and HCRF (e). Red and green show FN and TP regions respectively.

Table 1. Quantitative comparisons of the performance of different models. The first and second columns show the voxel-level classification results for the CRF using only the unary (CRFv(ϕ)) and the combination of the unary and pairwise potentials (CRFv(ϕ, φ)) respectively. The third column (HCRF) is the proposed hierarchical model. The forth and fifth columns show results using the CRF model found in [1], with (CRF[1]) and without post processing (CRF[1]$_{no\ post}$) respectively. The last column shows the classification result using an MRF model.

	CRFv(ϕ)	CRFv(ϕ,φ)	**HCRF**	CRF[1]	CRF[1]$_{no\ post}$	MRF
Sensitivity	0.98	0.98	**0.98**	0.88	0.92	0.81
Avg FPs	22.29	9.13	**1.55**	1.58	12.03	2.45
PPV	0.13	0.27	**0.66**	0.58	0.27	0.48

CRF level has decreased the average FP count by 59% (while maintaining high sensitivity) over using the unary potential alone. For the complete HCRF, incorporation of the higher order potential in the second level has decreased the average FPs by 83%. The low sensitivity rate in CRF [1] is due to missed lesions or the 1 or 2 lesion voxels filtered by post-processing. It should be noted that even though the average FP count is almost the same for the CRF [1] and HCRF, no post-processing is performed in HCRF to remove the small regions (and the sensitivity is much higher). As we see for $CRF[1]_{no\ post}$, the FP count increases drastically without any post-processing step. The Ising parameter for the MRF model was set to give an average FP count close to that of the HCRF model. Even with a higher FP count, the MRF has lower sensitivity as a result of oversmoothing.

4   Discussion

In this paper, we propose a new a Hierarchical CRF model to detect gad-enhancing lesions in brain MRI that embeds contextual information at multiple

levels. At the first level, a CRF model with up to pairwise potentials are used to detect the lesion candidates. In the second level, we group the similar lesion candidates according to features. For each group, a new CRF model embeding higher order potentials is learned, allowing for the integration of high level features (e.g. shape and symmetry) and boosting the discrimination power of our model. Our classifier is tested on 30 multi-centre clinical data set from RR MS patients with varying loads of gad-enhancing lesions. The experimental results show the advantage of the method, showing higher sensitivity rate while maintaining a very low false positive rate over other approaches. In the future, a full cross validation will be performed to precisely study the sensitivity of the parameters to different training sets. More sophisticated models for the higher order potential and investigation of the performance of the CRF classifier for other contexts are among the future works as well.

References

1. Karimaghaloo, Z., Shah, M., Francis, S., Arnold, D.L., Collins, D.L., Arbel, T.: Automatic detection of Gadolinium-enhancing multiple sclerosis lesions in brain MRI using conditional random fields. IEEE Trans. Med. Imag. 31(6), 1181–1194 (2012)
2. Miki, Y., Grossman, R., Udupa, J., Samarasekera, S., Buchem, M., Cooney, B., Kolson, D., Constantinescu, C., Pollack, S., Polansky, M., Mannon, L.: Computer-assisted quantitation of enhancing lesions in multiple sclerosis: correlation with clinical classification. Am. J. Neuroradiol. 18, 705–710 (1997)
3. Bedell, B., Narayana, P.: Automatic segmentation of Gadolinium-enhanced multiple sclerosis lesions. Magn. Reson. Med. 39, 935–940 (1998)
4. He, R., Narayana, P.: Automatic delineation of Gd enhancements on magnetic resonance images in multiple sclerosis. Med. Phys. 29, 1536–1546 (2002)
5. Datta, S., Sajja, R., He, R., Gupta, K., Wolinsky, S., Narayana, A.: Segmentation of gadolinium-enhanced lesions on MRI in multiple sclerosis. J. Magn. Reson. Imaging 25, 932–937 (2007)
6. Kohli, P., Ladicky, L., Torr, P.: Robust higher order potentials for enforcing label consistency. Int. J. Comput. Vision, 302–324 (2009)
7. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: CML, pp. 282–289 (2001)
8. Sutton, C., McCallum, A.: Piecewise pseudo-likelihood for efficient training of conditional random fields. In: ICML, pp. 863–870 (2007)
9. Boykov, Y., Kolmogorov, V.: An experimental comparison of min- cut/max-flow algorithms for energy minimization in vision. IEEE T. Pattern Anal., 1124–1137 (2004)
10. Nyul, L., Udupa, J.: On standardizing the MR image intensity scale. Magn. Reson. Med. 42, 1072–1081 (1999)

Simplified Labeling Process for Medical Image Segmentation

Mingchen Gao[1], Junzhou Huang[2], Xiaolei Huang[3],
Shaoting Zhang[1], and Dimitris N. Metaxas[1]

[1] CBIM Center, Rutgers University, Piscataway, NJ, 08554
[2] Department of Computer Science and Engineering,
University of Texas at Arlington, TX, 76019
[3] Computer Science and Enginnering Department, Lehigh University, PA, 18015

Abstract. Image segmentation plays a crucial role in many medical imaging applications by automatically locating the regions of interest. Typically supervised learning based segmentation methods require a large set of accurately labeled training data. However, thel labeling process is tedious, time consuming and sometimes not necessary. We propose a robust logistic regression algorithm to handle label outliers such that doctors do not need to waste time on precisely labeling images for training set. To validate its effectiveness and efficiency, we conduct carefully designed experiments on cervigram image segmentation while there exist label outliers. Experimental results show that the proposed robust logistic regression algorithms achieve superior performance compared to previous methods, which validates the benefits of the proposed algorithms.

1 Introduction

To assist doctors locate pathologies, automatic segmentation of different regions of medical images is very useful. Supervised learning based segmentation method, which use manually segmented training data as references, has superior performance. Those methods perform well as long as the feature space sufficiently distinguishes each label. They are relatively computationally efficient and not sensitive to parameters. Many popular learning methods have been applied to solve challenging medical problems, such as, but not limited to support vector machine (SVM) [1], neural network [2,3], conditional random field (CRF) [4], logistic regression (LR) [5].

One disadvantage of the learning-based segmentation is the requirement of manual interaction for obtaining training data. Manual segmentation by different people are subjective due to the lack of standard when performing the manual segmentation. In reality, the segmentation training data may not perfectly labeled. However, all the previous methods ignored the imperfection of training data. In such cases the training set is misleading, the guidance given by the labels may not be reliable. Consequently learning results may not totally reliable. On the other hand, the precisely labeling prcess is time consuming,

laborious and herpaps not necessary. In this paper, we take the advantage of learning methods handling label outliers and propose a simplified lableing process. The proposed simplified labeling process could significantly reduce doctors time and effort in labeling the training data. And the proposed learning methods would handle the labeling errors introduced by the new labeling process. While learning in the idealized setting has been thoroughly studied, learning in the presence of labeling errors has not been well explored. And nothing has been done for medical image segmentation with labeling error.

According to the practical consideration mentioned above, we are the first to systematically study logistic regression and sparse logistic regression method with outliers and apply it on cervigram image segmentation. Logistic regression (LR) is a classical method for classification and has been widely used in many applications. Sparse logistic regression, which is the logistic regression with $l1$-norm regularization, is a very effective method for classification on large scale data with high dimension. Both methods could be adopted in image segmentation.

In this paper, we propose a robust sparse logistic regression method to handle the classification problem with label outliers. In our method, a latent variable is introduced for the true correct labels, and then we estimate the probability of labels being flipped. Conjugate gradient method and Lassplore algorithm [6] are used respectively to minimize the loss function under the noise assumption. After optimization, the probability of the flipped labels is updated using the new classification estimation. The algorithm can quickly converge after several iterations.

To demonstrate the effectiveness of the proposed algorithm, we apply the proposed method on the task of automatically segmenting the biomarker AcetoWhite (AW) regions in an archive of $60,000$ images of the uterine cervix [7]. The most important observation in a cervigram image is the AW region, which is caused by whitening of potentially malignant regions of the cervix epitheliuem, following application of acetic acid to the cervix surface. Since the texture, size and location of AW regions have been shown to correlate with the pathologic grade of disease severity, accurate identification and segmentation of AW regions in cervigrams have significant implications for diagnosis and grading of cervical lesions. Carefully designed experiments on the cervigram images with label outliers demonstrate the superior performance of the proposed method and validate the benefits of the proposed algorithm in practical applications.

2   Related Work

In this section, we briefly review logistic regression and sparse logistic regression.

Logistic regression is a conditional probability of the label y, given a sample $x \in \mathbb{R}^n$, and a weight vector $w \in \mathbb{R}^n$:

$$p(y = 1 \mid x, w) = \sigma(w^T x) = \frac{1}{1 + e^{-w^T x}}. \tag{1}$$

Given a set of training examples $D = \{(x_1, \hat{y}_1), ..., (x_i, \hat{y}_i), ..., (x_m, \hat{y}_m)\}$, we want to learn a classifier $y = f(x)$. Here $x_i \in \mathbb{R}^n$ denotes a sample, and $\hat{y}_i$ is the corresponding observed class label. The likelihood function associated with these m samples is defined as

$$f(w) = log \prod_{i=1}^{m} p(\hat{y}_i \mid x_i, w)$$

$$= \sum_{i=1}^{m} \frac{\hat{y}_i + 1}{2} log(p(\hat{y}_i = 1 \mid x_i, w)) + \frac{1 - \hat{y}_i}{2} log(p(\hat{y}_i = -1 \mid x_i, w)),$$

which is a smooth and convex function. We can determine w by minimizing the logistic loss:

$$\hat{w} = \underset{w}{\operatorname{argmin}} f(w). \tag{2}$$

This is a smooth convex optimization problem. Regularization is usually used to avoid the overfitting problem. The l_2-norm regularization leads to a smooth and differentiable unconstrained convex optimization problem. Standard optimization algorithm such as Newton method and conjugate gradient method can be applied for solving such optimization problem [8,9].

The l_1-norm regularization is used to obtain a sparse model. However, the l_1-norm regularization term is non-differentiable. There are many algorithms proposed in the past for solving the l_1 regularized logistic regression [10,11]. Liu et al. proposed the Lassplore algorithm for solving large scale sparse logistic regression. They formulate the problem as the l_1-ball constrained logistical regression formulation, in which the objective function is continuously differentiable, and the problem domain set is closed and convex [6].

3   Proposed Algorithm

We propose a new simplified labeling process. Doctors would label a bounding box around the region of interest rather than precisely label the boundaries. Please see Figure 2 (a,b,c) as illustration of the proposed labeling process. Figure 2(a) is the image for segmentation. As we can see from (b), the groundtruth, the boundary of the region of interest in nontrivial and complicated. Our labeling process is illustration in (c), which would roughly include the region of interest in a bounding box. The simplified labeling process would significantly save time and effort of doctors in labeling the training data.

Considering the new labeling process, which introduces some labeling noise, none of the previous work has been able to handle label outliers. In this paper, we propose a robust logistic regression method to handle the classification problem on large scale data with label outliers.

3.1   Robust Logistic Regression

If there exists some label noise, Eq. 2 is no longer valid. We introduce a latent variable y_i, which represents the true label. The probability of the observed label $p(\hat{y}_i \mid x_i, w)$ is written as the following:

Table 1. Probabilistic relationship between observed label and true label

	Observed Label	
	-1	1
True -1	γ_{00}	γ_{01}
Label 1	γ_{10}	γ_{11}

$$
\begin{aligned}
S_i^k &\doteq p(\hat{y}_i = k \mid x_i, w) \\
&= \sum_{i=0}^{K} p(\hat{y}_i = k \mid y_i = j) p(y_i = j \mid x_i, w).
\end{aligned}
\tag{3}
$$

Here $p(\hat{y}_i = k \mid y_i = j)$ represents the probability that the label has flipped from the true label j to the observed label k. The probability relationship between observed label and true label is represented in Table 1. By expanding the terms in Eq. 3 we find that,

$$
\begin{aligned}
S_i^0 &= p(\hat{y}_i = -1 \mid x_i, w) \\
&= p(\hat{y}_i = -1 \mid y_i = -1) p(y_i = -1 \mid x_i, w) + p(\hat{y}_i = -1 \mid y_i = 1) p(y_i = 1 \mid x_i, w) \\
&= \gamma_{00}(1 - \sigma(w^T x)) + \gamma_{10}\sigma(w^T x) \\
S_i^1 &= p(\hat{y}_i = 1 \mid x_i, w) \\
&= p(\hat{y}_i = 1 \mid y_i = -1) p(y_i = -1 \mid x_i, w) + p(\hat{y}_i = 1 \mid y_i = 1) p(y_i = 1 \mid x_i, w) \\
&= \gamma_{01}(1 - \sigma(w^T x)) + \gamma_{11}\sigma(w^T x).
\end{aligned}
\tag{4}
$$

By substituteing Eq. 4 into Eq. 2, we find:

$$
f(w) = \sum_{i=1}^{m} \frac{\hat{y}_i + 1}{2} log(S_i^1) + \frac{1 - \hat{y}_i}{2} log(S_i^0).
\tag{5}
$$

This formulation remains smooth and differentiable. It can be solved using standard optimization algorithm such as conjugate gradient method [9,8].

3.2 Robust Sparse Logistic Regression

To handle outliers in sparse logistic regression, we need to optimize Eq. 5 subject to the l_1-ball constrain. We use the Lassplore algorithm to solve the sparse logistic regression problem [6]. The Lassplore algorithm is a first-order blackbox method that evaluates the function value and the gradient at each iteration. The gradient of Eq. 5 is the following:

$$
\nabla_w f(w) = \sum_{i=1}^{m} \frac{\hat{y}_i + 1}{2S_i^1} \nabla_w S_i^1 + \frac{1 - \hat{y}_i}{2S_i^0} \nabla_w S_i^0,
\tag{6}
$$

where $\nabla_w S_n^0$ and $\nabla_w S_n^1$ are:

$$\begin{aligned}
\nabla_w S_i^0 &= (\gamma_{10} - \gamma_{00})\sigma(w^T x_i)(1 - \sigma(w^T x_i)) \cdot x_i, \\
\nabla_w S_i^1 &= (\gamma_{11} - \gamma_{01})\sigma(w^T x_i)(1 - \sigma(w^T x_i)) \cdot x_i,
\end{aligned} \tag{7}$$

$\sigma(w^T x_n)$ is the conditional probability of $\hat{y}_i$ given a sample x_i.

An initial estimation of the label flipping probability can be given to start the optimization. After solving the Lassplore algorithm, label flipping probability is updated as the following:

$$\begin{aligned}
e_i &= \sigma(w^T x_i), \\
\gamma_{00} &= \frac{(e_i < 0.5)\&(\hat{y}_i = -1)}{e_i < 0.5}, \gamma_{01} = 1 - \gamma_{00}, \\
\gamma_{11} &= \frac{(e_i > 0.5)\&(\hat{y}_i = 1)}{e_i > 0.5}, \gamma_{10} = 1 - \gamma_{11},
\end{aligned} \tag{8}$$

To conclude, an iterative framework is proposed to update the label-flipping probability. First given an initialization of the label-flipping probability, the Lassplore method is used to minimize the loss function. Then, the label-flipping probability is estimated again using the new classification results. Using this scheme we iteratively update the label-flipping probability and solve the optimization problem.

4 Experimental Results and Discussion

We first use a synthetic dataset to illustrate the effectiveness of our algorithm. The examples are sampled from two multivariate normal distributions differing in mean and covariance. Here we focus on the non-uniform flipping of labels, where
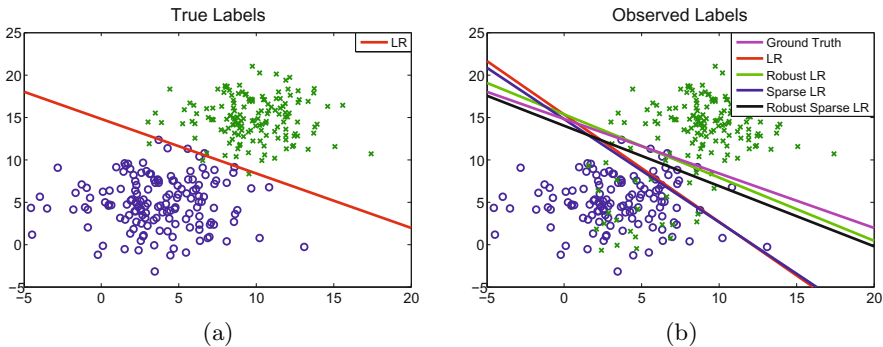
Fig. 1. (a) shows the samples with true labels and use Logistic Regression (LR) for separation. (b) shows the samples with observed labels, where about 20% labels from negative class are flipped to the positive class. Robust algorithms push the classification line to the true boundary line on true labels.

Table 2. Trained on the corrupted training labels, tested on true labels. The proposed methods show better performance.

Method	Sensitivity	Specificity	DSC
LR	0.9307 ± 0.0207	0.9987 ± 0.0027	0.9570 ± 0.0158
rLR	0.9920 ± 0.0040	0.9687 ± 0.0031	**0.9809 ± 0.0013**
SLR	0.9153 ± 0.0248	0.9987 ± 0.0027	0.9508 ± 0.0158
rSLR	0.9840 ± 0.0033	0.9807 ± 0.0020	**0.9833 ± 0.0015**

(a) Image for segmentation   (b) The groundtruth   (c) Roughly labeled data

(d) LR using groundtruth   (e) LR   (f) rLR

(g) SLR   (h) rSLR
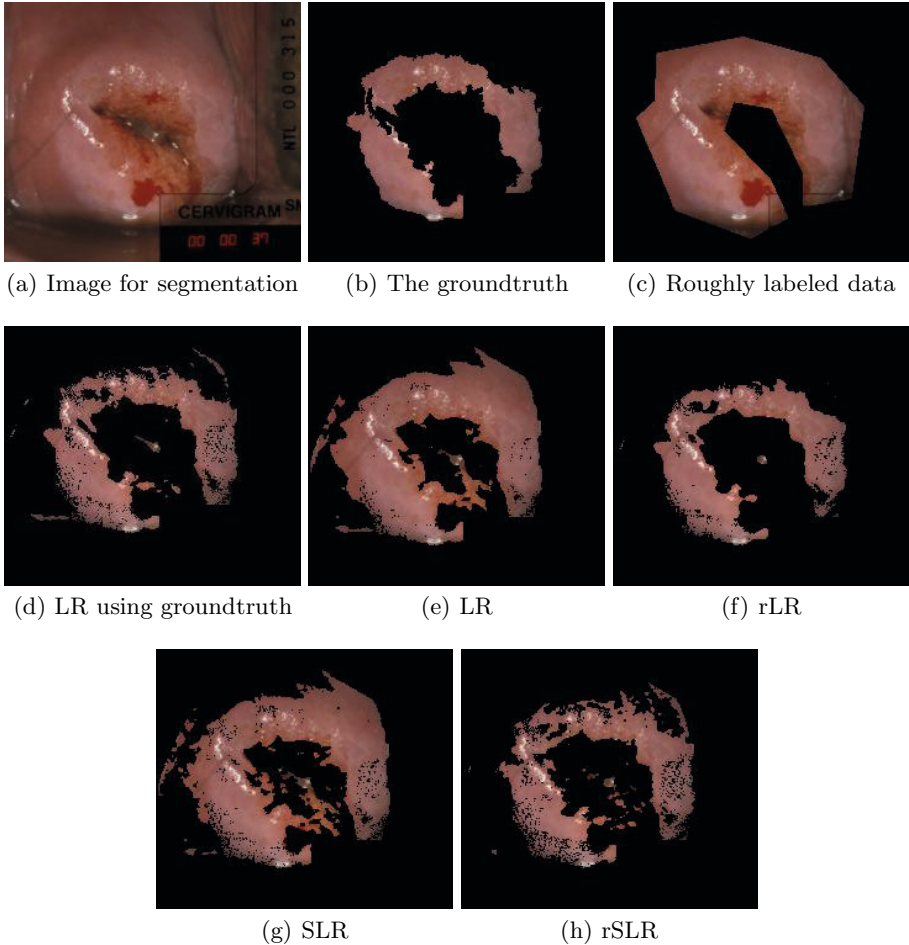
Fig. 2. (d) uses the groundtruth and LR in training for comparison. The simplifiled labeling data, as shown in (c) are used for training for robust algorithm. (f) and (h) include lots of negative points since we use some of the negative points in training. Use the robust algorithms, the result expelled many negative points.

Table 3. Testing errors on segmenting the cervigram data

Method	Sensitivity	Specificity	DSC
LR using groundtruth	0.8351	0.8820	0.8157
LR	0.9486	0.5660	0.6950
rLR	0.8175	0.8825	**0.8058**
SLR	0.9354	0.5931	0.7012
rSLR	0.7121	0.8986	**0.7517**

there are negative data being labeled as positive ones. This setting is to imitating the proposed labeling process in medical images. All the data are training on the corrupted data and then tested on the true labels. We quantitatively evaluate the classification results in Table 2. Compared to logistic regression (LR) and sparse logistic regression (SLR) methods, both robust logistic regression (rLR) and robust sparse logistic regression (rSLR) perform better on sensitivity but worse on specificity. Dice similarity coefficient(DSC), which is defined as $(2 * \text{true positive})/(2 * \text{true positive} + \text{false positive} + \text{false negative})$, measures the consistency of results and groundtruth. DSC is a more comprehensive measurement of the coinicdence of the classification result with the groundtruth. The proposed methods perform better on this measurement.

We apply the proposed algorithm to classify pixels in optical cervigram images into two classes, namely normal and abnormal tissues while not given accurate segmentation training labels. The experiment is designed as the following: We have experts precisely labeled the segmentation, which is considered as groundtruth. We also have experts roughly labeled the segmentation, using our proposed simplified labeling method. The roughly labeled data is used for training, and then we test our segmentation results against the precisely labeld segmentation. Our proposed labeling process and segmentation results would achieve competitive results compared to using the the groundtruth for training. In our experiments, HSV feature is chosen for classification because it achieves the best performance.

Table 3 shows the results of different classifiers measured by sensitivity, specificity and DSC. Our method consistently achieves significantly better performance in terms of DSC. The results from our method also compares favorably with other state-of-the-art methods in this application [12].

5   Conclusion

This paper proposed a novel method on handling labeling noise. The proposed method has been demonstrated on cervigram segmentation problem. Using a bounding box around the region of interest as labels other than accurate labels, we can still get comparable results. In the future, we want to extend our work using group sparsity [13].

References

1. Wang, S., Zhu, W., Liang, Z.-P.: Shape deformation: SVM regression and application to medical image segmentation. In: ICCV, vol. 2, pp. 209 –216 (2001)
2. Shen, S., Sandham, W., Granat, M., Sterr, A.: MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization. IEEE Transactions on Information Technology in Biomedicine, 9(3), 459–467 (2005)
3. Cheng, K.S., Lin, J.S., Mao, C.W.: The application of competitive hopfield neural network to medical image segmentation. IEEE Transactions on Medical Imaging 15(4), 560–567 (1996)
4. Lee, C.-H., Schmidt, M., Murtha, A., Bistritz, A., Sander, J., Greiner, R.: Segmenting Brain Tumors with Conditional Random Fields and Support Vector Machines. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 469–478. Springer, Heidelberg (2005)
5. Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M., Murtha, A.: 3D variational brain tumor segmentation using a high dimensional feature set. In: ICCV, pp. 1–8 (2007)
6. Liu, J., Chen, J., Ye, J.: Large-scale sparse logistic regression. In: KDD, pp. 547–556 (2009)
7. Gordon, S., Zimmerman, G., Long, R., Antani, S., Jeronimo, J., Greenspan, H.: Content analysis of uterine cervix images: Initial steps towards content based indexing and retrieval of cervigrams (2006)
8. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
9. Minka, T.P.: A comparison of numerical optimizers for logistic regression. Technical report (2003)
10. Koh, K., Kim, S.-J., Boyd, S.P.: An interior-point method for large-scale l_1 regularized logistic regression. Journal of Machine Learning Research 8, 1519–1555 (2007)
11. Ng, A.Y.: Feature selection, l1 vs. l2 regularization, and rotational invariance. In: ICML, pp. 78–86 (2004)
12. Zhang, S., Huang, J., Metaxas, D.N., Wang, W., Huang, X.: Discriminative sparse representations for cervigram image segmentation. In: ISBI, pp. 133–136 (2010)
13. Huang, J., Huang, X., Metaxas, D.N.: Learning with dynamic group sparsity. In: ICCV, pp. 64–71 (2009)

Liver Segmentation Approach Using Graph Cuts and Iteratively Estimated Shape and Intensity Constrains

Ahmed Afifi[1] and Toshiya Nakaguchi[2]

[1] Faculty of Computers and Information, Menoufia University, Egypt
[2] Graduate School of Engineering, Chiba University, Japan

Abstract. In this paper, we present a liver segmentation approach. In which, the relation between neighboring slices in CT images is utilized to estimate shape and statistical information of the liver. This information is then integrated with the graph cuts algorithm to segment the liver in each CT slice. This approach does not require prior models construction, and it uses single phase CT images; even so, it is talented to deal with complex shape and intensity variations. Moreover, it eliminates the burdens associated with model construction like data collection, manual segmentation, registration, and landmark correspondence. In contrast, it requires a low user interaction to determine the liver landmarks on a single CT slice only. The proposed approach has been evaluated on 10 CT images with several liver abnormalities, including tumors and cysts, and it achieved high average scores of 81.7 using MICCAI-2007 Grand Challenge scoring system. Compared to contemporary approaches, our approach requires significantly less interaction and processing time.

1 Introduction

In liver CAD systems, the liver segmentation is the first and essential process, and its accuracy is of special significance. However, this process is difficult because of low contrast between the liver and surrounding tissues , great differences in liver shape and intensity , and the existence of liver abnormalities. In literature, there are many attempts to solve the liver segmentation problem and various approaches have been proposed, including intensity or texture based approaches, deformable and statistical model-based approaches, and probabilistic atlases based approaches. Survey and comparison of different liver segmentation approaches have been presented in [1,2].

In the intensity based approaches, one or multiple intensity thresholds, region growing, or watershed methods are applied to extract an initial binary volume which consequently refined using morphological filters or knowledge-based approaches. Recent approaches of this category have been proposed in [3,4], and by Beck and Aurich in [2]. In the deformable model-based approaches, an initial contour or surface is deformed to minimize a predefined energy function. In [5,6], deformable models have been coupled with shape models and intensity thresholding to perform liver segmentation. Additionally, Gradient vector flow (GVF)

active contour [7] has been utilized for liver segmentation by R. S. Alomari *et al.* in [8] and by Chi *et al.* in [2]. The implicit deformable models, also called implicit active contours or level sets [9], have been utilized for liver segmentation as well. The statistical models have been received high interest from the investigators of liver segmentation approaches. They construct linear or non-linear models to represent the variation in liver shape and appearance like the approaches presented in [10,11,12]. In addition to statistical shape models, probabilistic atlases have been integrated into different liver segmentation approaches [13].

Despite this prosperous literature, we can conclude that the intensity- and deformable-based approaches were highly affected by the liver abnormalities. The statistical model- and probabilistic atlas-based approaches could enhance the results; however, they added a burden of model construction and matching. In this paper therefore, we present a knowledge-based liver segmentation approach. In this approach, we benefit from the high correlation between consequent slices of the same patient to define the shape constrains, and to estimate the statistical parameters of the liver and non-liver tissues. For initialization, the user segment one slice in the volume to define these constrains, and consequently they automatically updated from the nearby slices. A graph cuts algorithm based on the defined constrains is applied in a slice-by-slice manner to automatically segment the whole volume. Additionally, to reduce the computational time, we build the graph in a narrow band area defined from the adjacent slice. This proposed approach share the concept of constrains propagation with the method of *Lee et al.* [14]. However, the segmentation is performed from large to small liver cross sections which increases the ability of capturing separated and damaged liver parts. Moreover, shape and intensity constrains are integrated directly into the graph cuts segmentation algorithm and they are updated based on the segmentation results of the slices that have been processed so far.

The rest of this paper is organized as follows: in Sect. 2, the proposed approach is described. The evaluation results of the proposed approach are presented and discussed in Sect. 3. Finally, the paper is concluded in Sect. 4.

2   Proposed Segmentation Approach

The proposed segmentation approach mimics the human methodology in determining the boundary of liver. In this methodology, the correspondence between adjacent slices in CT image helps in alleviating the ambiguity of the liver boundary and in detecting the liver abnormalities. The whole procedure of the proposed approach is as follows:

Step-1: Performing image normalization in soft tissue window and then applying nonlinear diffusion filter to each slice.
Step-2: Selecting one slice containing nearly the largest liver cross section as the start slice and then define the liver object on it.
Step-3: Estimating initial shape, intensity, and graph cuts constrains from the start slice.

for all lower slices, starting from the start slice to the last one. **do**

    Step-4: Define a narrow band around the liver object.

    Step-5: Performing slice segmentation using shape-based graph cuts algorithm.

    Step-6: Adding the segmentation results of this slice to the output volume.

    Step-7: Updating shape, intensity, and graph cuts constrains according to the segmentation results of the current slice.

end for

for all upper slices, starting from the start slice to the first one. **do**

    Repeat Step-4 and Step-5.

    if the segmented object contains multiple parts **then**

        Step-8: Selecting the left most one as the liver object.

    end if

    Repeat Step-6 and Step-7.

end for

Step-9: Applying the postprocessing procedure to the output volume.

2.1 Preprocessing

The first aim of this process is to map the raw CT data encoded in either twelve or sixteen bits to gray scale data encoded in eight bits. The mapping or normalization is performed in a soft tissue window determined by selecting the lower (Lo) and upper (Hi) bounds of the right distribution in the histogram of the raw CT data. This mapping is performed according to (1).

$$I_g(x, y) = \begin{cases} 0 & \text{if} \quad I_o(x, y) \leq Lo \\ \frac{255(I_o(x,y) - Lo)}{Hi - Lo} & \text{if } Lo \leq I_o(x, y) \leq Hi, \\ 255 & \text{if} \quad I_o(x, y) \geq Hi \end{cases} \tag{1}$$

where, I_o is the raw CT image, I_g is the produced gray level image.

After mapping the whole CT volume to a gray scale volume, a nonlinear diffusion filter [15] is applied to each 2D slice in the volume to reduce the noise and increase the liver homogeneity.

2.2 Estimation of the Shape and Intensity Constrains

The shape constrains are applied as a prior probability of the liver location, and the intensity constrains are defined as the probability of the liver intensity model at each pixel. These constrains are automatically determined for each slice according to the segmented liver in the previous slice. The estimation process is performed according to the following procedure.

1. Define; the binary image of liver segmentation in the start slice as $Temp_{str}$, the binary liver object in this slice as $objectinTemp_{str}$, the binary image of liver segmentation in the previous slice as $Temp_{prv}$, the binary liver object in this slice as $objectinTemp_{prv}$, the pixels belonging to the liver in the previous
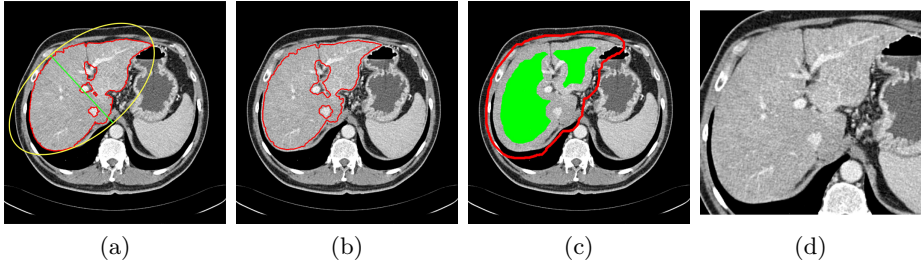
Fig. 1. Constrains estimation, (a) sample previous slice (liver contour in red and the minor axis in green), (b) the contour of the estimated shape template shown on the current slice, (c) the estimated constrains for graph cut (object in green and background in red), and (d) the slice after applying the narrow band constrain

slice as $Liver_{prv}$, and the pixels not belonging to the liver in the pervious slice as $non - Liver_{prv}$.

2. Determine the minor axis of the ellipse that fit the object in $Temp_{prv}$ and denote it as m_{ax} (Fig. 1a).

3. Erode the $Temp_{prv}$ with a disk structuring element of radius round$(0.02 \times m_{ax})$ and considering the result as the shape template of the current slice (Fig. 1b). This erosion value has been decided after studying the average change of the minor liver axes in different cases.

4. If Area($objectinTemp_{prv}$) $\geq 0.1 \times$ Area($objectinTemp_{str}$),calculate the histogram of $Liver_{prv}$ and $non - Liver_{prv}$ as the intensity model; else, use the previously used intensity model.

5. Erode $Temp_{prv}$ with a disk structuring element of radius round$(0.1 \times m_{ax})$ and the result is considered as the object hard constrains in the graph cuts algorithm (Fig. 1c).

6. Dilate the $Temp_{prv}$ with a disk structuring element of radius $max(2, \text{round}(0.1 \times m_{ax}))$. Then, the edge of the resulting binary template is determined and dilated with a disk structuring element of radius 1. The result of this step is considered as the background hard constrains in the graph cuts algorithm (Fig. 1c).

7. Define a narrow band window surrounding the liver object as the smallest rectangle fitting the dilated object calculated in Step 5 (Fig. 1d).

2.3   Segmentation Using Graph Cuts

The aim of this process is to find a labeling $A = \{A_1, A_2, , A_p, , A_{|P|}\}$ which minimize the the total energy function considering the estimated constrains as in (2).

$$E_T(A) = (1 - \lambda)R_D(A) + \lambda R_s(A) + \mu B(A), \qquad (2)$$

where, μ determines the relative importance of the boundary term ,$B(A)$, versus the regional term and λ determines the relative importance of the data penalty, $R_D(A)$, versus the shape penalty ,$R_s(A)$. The data penalty reflects on how the

intensity of a pixel fits into the intensity model of the object (liver) and background (non liver tissues). The shape penalty is encoded as the prior probability of a pixel to be inside or outside the liver object. The data, shape, and boundary penalties are calculated as in (3), (4), and (5), respectively.

$$
R_D\left(A_p\right) =
\begin{cases}
\dfrac{\log\left(pr\left(I_p \in \text{"}obj\text{"}\right)\right)}{\log\left(pr\left(I_p \in \text{"}obj\text{"}\right)\right)+\log\left(pr\left(I_p \in \text{"}bkg\text{"}\right)\right)} & \text{if } A_p = 1 \\[3mm]
\dfrac{\log\left(pr\left(I_p \in \text{"}bkg\text{"}\right)\right)}{\log\left(pr\left(I_p \in \text{"}obj\text{"}\right)\right)+\log\left(pr\left(I_p \in \text{"}bkg\text{"}\right)\right)} & \text{if } A_p = 0
\end{cases}
\tag{3}
$$

$$
R_s\left(A_p\right) =
\begin{cases}
1 - shape_{temp} & \text{if } A_p = 1 \\
shape_{temp} & \text{if } A_p = 0
\end{cases}
\tag{4}
$$

$$
B_{pq} = e^{-\frac{\left|I_p - I_q\right|^2}{2\sigma^2}} \times \frac{1}{d(p,q)},
\tag{5}
$$

where, $shape_{temp}$ is the estimated shape template of the objected (liver) in the current slice, I_p is the intensity value of a pixel p, $pr\left(I_p \in \text{"}obj\text{"}(\text{"}bkg\text{"})\right)$ is the probability of p to be an object("obj") or background ("bkg") pixel, and $d(p,q)$ is the Euclidian distance between pixels p and q.

This total energy function can be minimized efficiently using the graph cuts algorithm [16]. To achieve this goal, a graph with cut cost equaling the value of $E_T(A)$ is constructed using the edge weights defined in (6), (7), and (8). Furthermore, the hard constrains defined in Sect. 2.2 are implemented via infinity cost edges.

$$
w_{sp} =
\begin{cases}
\infty & \text{if } p \in \text{"}obj\text{"} \\
0 & \text{if } p \in \text{"}bkg\text{"} \\
(1-\lambda)R_D\left(A_p = 0\right) + \lambda R_s\left(A_p = 0\right) & otherwise
\end{cases}
\tag{6}
$$

$$
w_{pt} =
\begin{cases}
0 & \text{if } p \in \text{"}obj\text{"} \\
\infty & \text{if } p \in \text{"}bkg\text{"} \\
(1-\lambda)R_D\left(A_p = 1\right) + \lambda R_s\left(A_p = 1\right) & otherwise
\end{cases}
\tag{7}
$$

$$
w_{pq} = B_{pq}(p,q),
\tag{8}
$$

w_{sp}, w_{pt} are the weight of the links to terminal nods, and w_{pq} is the weight of the link between two adjacent pixels.

2.4   Postprocessing

In this process, any tissue surrounded by the segmented liver tissue is added to the final segmentation which smoothed using a 3D filter. To achieve this goal the following procedure has been applied.

1. Perform hole filling to each 2D slice.
2. Perform binary image closing to the whole 3D volume using a ball structuring element of radius 3.
3. Perform hole filling to each 2D slice again.
4. Smooth the final volume by applying a binary median filter of $3 \times 3 \times 3$ size..

3   Results and Discussion

Data Sets: The data set used for evaluation is the MICCAI2007 grand challenge test set [17]. This test set contains 10 CT images acquired using variety of CT scanners. In some cases, the entire anatomy is rotated around the z-axes. Most images in this data set have liver abnormalities, including tumors, metastasis, and cysts of different sizes.

Parameter Setting: All parameters have been adjusted using 5 CT images having different characteristics; 3 from MICCAI2007 training data set and 2 from a local data set. Graph cuts parameter μ was set to 2. The parameter λ was set to 0.2. The parameter σ in the boundary term was dynamically selected from each slice as the average absolute intensity difference between the neighboring pixels $(\sigma = \frac{1}{|P|} \sum_{p \in P, q \in N_p} |I_p - I_q|)$.

Evaluation Metrics: The proposed approach has been evaluated using the scoring system of MICCAI-2007 Grand Challenge workshop [2] which includes five metrics; Volumetric Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASD), Root Mean Square Symmetric Surface Distance (RMD), Root Mean Square Symmetric Surface Distance (RMD), and Maximum Symmetric Surface Distance (MSD). Moreover, the final precision score has been calculated according to the method presented by Heimann et. al. [2].

3.1   Experiments on Clinical Data

The segmentation approach has been implemented using Matlab environment on Windows-based personal computer with a Corei7(2.8GHz) processor and 6GB of memory. The evaluation results of the segmentation approach which calculated by the committee of the "3D Segmentation in the Clinic: A Grand Challenge" workshop of MICCAI2007 are shown in Table 1.

Comparative results of the proposed approach, the best automatic method (*Kainmüller et al.*) and all interactive methods reported by T. Heimann et. al. in [2] are shown in Table 2. As in [2], All approaches has been classified according to the time required for interaction. Less than 1 min was regarded as low interaction, less than 5 min as medium interaction, and more than 5 min as high interaction. Referring Table 2, the proposed approach share the best position with Beichel *et. al.* MBR. Additionally by referring the recent results on sliver07.org database [17], the proposed approach is in the third place of all methods. However, the proposed approach is significantly faster, requires less amount of interaction, and does not require extensive manual refinement. The automatic method of Kainmüller *et. al.* achieved this results by using an extensive training set of 112 liver shapes to build a statistical shape model (SSM) consists of around 7.000 landmarks. The total score of the same method was 73 when the number of training shapes used to build the SSM was 43 [10].

Since the shape and intensity constrains are estimated in a case-specific manner, the proposed approach is robust for liver shape variations and existence of liver abnormalities. Fig. 2 show that, the proposed approach can efficiently

Table 1. Evaluation results of the proposed approach

Case	VOE		RVD		ASD		RMD		MSD		Total Score	Time (sec.)	
	[%]	Score	[%]	Score	[mm]	Score	[mm]	Score	[mm]	Score		Initial	Total
#1	5.2	80	2.4	87	0.7	82	1.4	81	14.7	81	82	35	221
#2	5.9	77	5.0	74	0.8	80	1.7	76	19.4	74	76	40	223
#3	3.9	85	2.2	88	0.7	83	1.1	84	14.0	82	84	37	218
#4	5.0	80	2.5	86	0.7	81	1.4	80	10.4	86	83	36	122
#5	6.1	76	1.2	94	1.0	76	1.9	74	21.5	72	78	36	118
#6	5.8	78	0.7	96	0.8	79	1.8	74	20.1	74	80	37	204
#7	3.8	85	1.5	92	0.5	87	1.2	84	16.0	79	85	38	170
#8	6.2	76	1.1	94	1.0	75	2.3	68	22.2	71	77	35	113
#9	4.2	84	1.2	94	0.5	87	1.2	83	16.0	79	85	37	284
#10	4.5	82	0.5	98	0.6	86	1.2	84	11.5	85	87	36	108
Average	5	80	1.8	90	0.7	82	1.5	79	16.6	78	81.7	36.7	178.1
Std. Dev.	0.9	3.6	1.3	7.0	0.18	4.3	0.4	5.5	4.1	5.3	3.8	1.5	60.8

Table 2. Comparative results of the proposed segmentation approach

Method	VOE		RVD		ASD		RMD		MSD		Final	Runtime
	[%]	Score	[%]	Score	[mm]	Score	[mm]	Score	[mm]	Score	Score	[min]
Beichel *et. al.* MBR(high)	5.2	80	1.0	91	0.8	80	1.4	80	15.7	79	82	36
Proposed approach(low)	**5**	**80**	**1.8**	**90**	**0.7**	**82**	**1.5**	**79**	**16.6**	**78**	**82**	**3**
Kainmüller *et. al.*(Automatic)	6.1	76	−2.9	85	0.9	76	1.9	74	18.7	75	77	15
Beck and Aurich(high)	6.6	74	1.8	88	1.0	74	1.9	73	18.5	76	77	7
Dawant *et. al.*(med)	7.2	72	2.5	86	1.1	73	1.9	74	17.1	77	76	20
Second rater	6.4	75	4.7	75	1.0	75	1.8	75	19.3	75	75	
Lee *et. al.*(low)	6.9	73	1.3	88	1.1	73	2.1	71	21.3	72	75	7
Beichel *et. al.* CBR(med)	6.5	74	1.1	90	1.1	72	2.5	66	23.4	69	74	31
Wimmer *et. al.*(med)	8.1	68	6.1	68	1.3	67	2.2	69	18.7	75	69	4 − 7
Slagmolen *et. al.*(med)	10.4	59	3.7	70	2.0	50	5.0	34	40.5	47	52	60
Beichel *et. al.*(low)	14.3	48	3.1	62	3.6	34	7.9	24	49.2	38	41	30

(a)   (b)   (c)   (d)
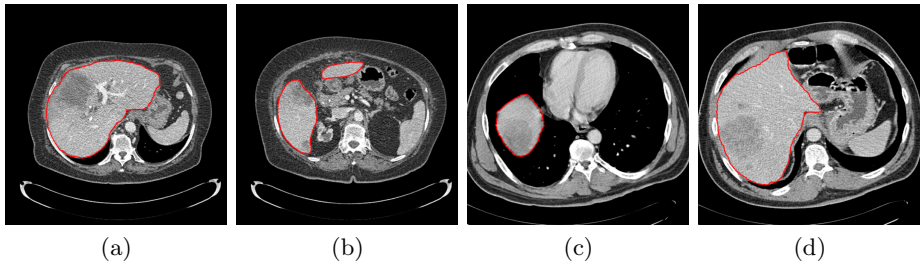
Fig. 2. Segmentation results of cases containing large and dense liver tumors

extract the liver in different cases containing large and dense tumors. Referring Table 1, the average performance of the proposed approach (81.7) cab be regarded as closer to the reference manual segmentation than the human performance (75) [2]. Small deviation of these scores shows the ability of the proposed approach to deal with extreme cases as well as easy and moderate cases. The processing time required to segment a CT volume ranges from 2 – 5 minutes and it is significantly less than the manual or other conventional segmentation methods. In general, the proposed approach can efficiently utilize the anatomical knowledge of the liver to achieve accurate segmentation results.

4   Conclusion

In this work, we proposed a novel shape-based approach for liver segmentation in portal-venous CT images using a case-specific knowledge. In which, the relation between consequent slices of the same image is exploited to estimate the shape and intensity information of the liver. Then, this information is integrated into the graph cuts algorithm to segment the whole CT image. Unlike the other shape-based segmentation approaches which use training data to build a statistical model, the proposed technique does not require prior model construction. Accordingly, it is not restricted to the trained model, and it can be applied when there is no training data available. The evaluation results demonstrated the high precision of the proposed approach. It efficiently estimates the liver boundary even with the existence of large and dense liver abnormalities. The utilization of a case-specific knowledge increases the ability of the proposed approach to deal with difficult and atypical liver shapes. Additionally, it removes the burden of model construction and matching. A low processing time required by the proposed approach makes it suitable for clinical application.

References

1. Campadelli, P., Casiraghi, E., Esposito, A.: Liver segmentation from computed tomography scans: A survey and a new algorithm. Artificial Intelligence in Medicine 45, 185–196 (2009)
2. Heimann, T., Ginneken, B.V., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P.M.M., Chi, Y., Córdova, A., Dawant, B.M., Fidrich, M., Furst, J.D., Furukawa, D., Grenacher, L., Hornegger, J., Kainmüller, D., Kitney, R.I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.P., Németh, G., Raicu, D.S., Rau, A.M., van Rikxoort, E.M., Rousson, M., Ruskó, L., Saddi, K.A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J.M., Wimmer, A., Wolf, I.: Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. IEEE Trans. Med. Imag. 28(8), 1251–1265 (2009)
3. Rusko, L., Bekes, G., Fidrich, M.: Automatic segmentation of the liver from multi- and single-phase contrast-enhanced CT images. Medical Image Analysis 13, 871–882 (2009)
4. Foruzana, A.H., Zoroofia, R.A., Horib, M., Satoc, Y.: Liver segmentation by intensity analysis and anatomical information in multi-slice CT images. International Journal of CARS 4(3), 287–297 (2009)
5. Tibamoso, G., Rueda, A.: Semi-automatic Liver Segmentation From Computed Tomography (CT) Scans based on Deformable Surfaces. SLIVER07 Results (October 2009), http://sliver07.isi.uu.nl/results/20091022201318/description.pdf
6. Gao, J., Kosaka, A., Kak, A.: A Deformable Model for Automatic CT Liver Extraction. Academic Radiology 12(9), 1178–1189 (2005)
7. Xu, C., Prince, J.: Snakes, shapes, and gradient vector flow. IEEE Trans. Image Process. 7(3), 359–369 (1998)

8. Alomari, R.S., Kompalli, S., Chaudhary, V.: Segmentation of the Liver from Abdominal CT Using Markov Random Field model and GVF Snakes. In: Proc. International Conference on Complex, Intelligent and Software Intensive Systems, pp. 293–298 (2008)

9. Sethian, J.A.: Level Set Methods and Fast Marching Methods, 2nd edn., pp. 1–74. Cambridge University Press (1996)

10. Kainmuller, D., Lange, T., Lamecker, H.: Shape Constrained Automatic Segmentation of the Liver based on a Heuristic Intensity Model. In: Proc. MICCAI Workshop on 3-D Segmentation in Clinic: A grand Challenge, pp. 109–116 (2007)

11. Afifi, A., Nakaguchi, T., Tsumura, N., Miyake, Y.: A Model Optimization Approach to the Automatic Segmentation of Medical Images. IEICE Trans. on Information and Systems E93-D(4), 882–889 (2010)

12. Linguraru, M.G., Pura, J.A., Chowdhury, A.S., Summers, R.M.: Multi-organ Segmentation from Multi-phase Abdominal CT via 4D Graphs Using Enhancement, Shape and Location Optimization. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 89–96. Springer, Heidelberg (2010)

13. Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y.W., Nakamura, H., Sato, Y.: Automated Segmentation of the Liver from 3D CT Images Using Probabilistic Atlas and Multilevel Statistical Shape Model. Academic Radiology 15(11), 1390–1399 (2008)

14. Lee, J., Kim, N., Lee, H., Seo, J.B., Won, H.J., Shin, Y.M., Shin, Y.G., Kim, S.-H.: Efficient liver segmentation using a level-set method with optimal detection of the initial liver boundary from level-set speed images. Computer Methods and Programs in Biomedicine 88, 26–38 (2007)

15. Weickert, J., Romeny, B.M., Viergever, M.A.: Efficient and Reliable Schemes for Nonlinear Diffusion Filtering. IEEE Trans. Image Process. 7(3), 398–410 (1998)

16. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Anal. Mach. Intell. 23(11), 1222–1239 (2001)

17. Heimann, T., Ginneken, B.V., Styner, M.A.: Segmentation of the Liver 2007 (SLIVER07), http://sliver07.isi.uu.nl/ (last visted: June 10, 2012)

Multi-Object Geodesic Active Contours (MOGAC)

Blake C. Lucas[1,2], Michael Kazhdan[2], and Russell H. Taylor[2]

[1] Johns Hopkins Applied Physics Laboratory, Laurel, MD, USA
[2] Johns Hopkins University, Baltimore, MD, USA
{blake,misha}@cs.jhu.edu, rht@jhu.edu

Abstract. An emerging topic is to build image segmentation systems that can segment hundreds to thousands of objects (i.e. cell segmentation \ tracking, full brain parcellation, full body segmentation, etc.). Multi-object Level Set Methods (MLSM) perform this task with the benefit of sub-pixel precision. However, current implementations of MLSM are not as computationally or memory efficient as their region growing and graph cut counterparts which lack sub-pixel precision. To address this performance gap, we present a novel parallel implementation of MLSM that leverages the sparse properties of the algorithm to minimize its memory footprint for multiple objects. The new method, Multi-Object Geodesic Active Contours (MOGAC), can represent N objects with just two functions: a label mask image and unsigned distance field. The time complexity of the algorithm is shown to be $O((M^d)/P)$ for M^d pixels and P processing units in dimension d={2,3}, independent of the number of objects. Results are presented for 2D and 3D image segmentation problems.

Keywords: active contours, segmentation, level set, parallel.

1   Introduction

The Level Set Method (LSM) [1, 2] is popular in computer vision systems for segmenting images [3]. LSM solves PDEs to produce image segmentations with sub-pixel accuracy. The multi-object version is capable of segmenting adjacent structures without gaps or overlaps [4-10]. However, current implementations of the Multi-object Level Set Method (MLSM) are slow and require a large memory footprint compared to their region growing [11] and graph cut counterparts [12], which lack sub-pixel accuracy. A modern challenge is to develop MLSM implementations that have competitive computational and memory efficiency with region growing and graph cut methods.

Several methods have been proposed for segmenting N objects with N level set functions [4, 6, 8] that are stored as images. Storage of these level set images is intractable for tasks such as cell tracking in microscopy images [13] where there are potentially hundreds to thousands of objects. The Multi-phase LSM [7] reduces the number of level sets to $log_2(N)$, and the Multi-compartment LSM [5, 9] reduces the number of functions to 4 in 2D and 6 in 3D. Even with these advancements, some segmentation tasks are still intractable because the time complexity for existing MLSMs is dependent on the number of objects. By comparison, region growing

techniques require only one function to represent N objects, and the computation time can be independent of the number of objects.

All previous MLSMs use implementations of the narrow-band method [1], which requires periodic re-initialization of the signed distance field. One notable exception is the work by Lie et al. [10], but their method does not have sub-pixel precision. The fast-marching method [1] for distance field re-initialization is a computational bottleneck whose complexity is $O(M^d log M)$ for M^d pixels in dimension $d \in \{2,3\}$. A more efficient approach is to use the sparse-field method [14] that has $O(M^{d-1})$ time complexity. The sparse-field LSM stores only the minimum narrow-band needed for finite difference calculations and maintains an approximation to the signed distance field at every time step. The sparse-field LSM is competitive with region growing methods and will be extended in this work to create a new MLSM.

To develop LSMs that run at faster, real-time speeds, implementations must leverage parallelism now abundant on modern GPUs and CPUs. There has been work on parallel implementations of single object LSMs [15] that achieve substantial speed-up. Memory consumption is a concern for parallel implementations because GPUs generally have access to less memory than CPUs; and at real-time speeds, memory latency and access patterns become a major performance concern. These concerns complicate development of a parallel Multi-object Level Set Method, which has yet to be proposed in literature.

The contributions of this work are two-fold. First, we describe how to represent N level sets with only a label mask and distance field. Second, we describe a parallel algorithm for segmenting N objects in 2D and 3D. Properties of the sparse-field LSM are leveraged to represent and evolve N level sets with the "label mask + distance field" data structure. Results are presented for 2D/3D segmentation of multiple, potentially overlapping, objects. The scalability of the algorithm is analyzed, and the computational complexity of the algorithm is discussed and juxtaposed with other MLSM implementations. We regard parallelism as a necessary consideration when proposing a new algorithm because computing hardware is becoming more parallel as opposed to faster. The algorithm is implemented in OpenCL and runs on the GPU, but the focus of this work will be on the algorithm, not GPU performance. Source code is available at http://code.google.com/p/imagesci/.

2   Method

Representation. The following LSM, which we refer to as Multi-Object Geodesic Active Contours (MOGAC), segments a gray level image $I: \Omega \mapsto \mathcal{R}$ with domain $\Omega \subset \mathcal{R}^d$ in dimension $d \in \{2,3\}$ into at most N object regions represented by signed distance fields $\varphi_n: \Omega \mapsto \mathcal{R}$ for labeled regions $n \in \mathcal{L} = \{1, \cdots, N\}$. The segmentation is compressed into a label function $\chi: \Omega \mapsto \mathcal{L}$ and unsigned distance field $\psi: \Omega \mapsto \mathcal{R}$ via $\psi(x) = \min_n |\varphi_n(x)|$. An approximation to $\varphi_n(x)$ is computable from $\chi(x)$ and $\psi(x)$ at the boundary of region n:

$$\Lambda_n = \{x | \exists y \epsilon \mathcal{N}(x) \text{ s.t. } \sigma_n(x) \neq \sigma_n(y)\} \qquad (1)$$

where $\mathcal{N}(x)$ is the connected neighborhood of pixel x and the sign $\sigma_n(x)$ is indicated by,

$$\sigma_n(x) = \begin{cases} -1 & \chi(x) = n \\ 1 & otherwise \end{cases}.$$

$$(2)$$

The partially reconstructed level set $\tilde{\varphi}_n : \Lambda_n \mapsto \mathcal{R}$ is given by $\tilde{\varphi}_n(x) = \sigma_n(x)\,\psi(x)$. $\tilde{\varphi}_n$ accurately measures the signed distance at the boundary of two objects. At the shared boundary of more than two objects, this measurement is approximate. $\tilde{\varphi}_n$ provides enough information to extract an iso-surface with marching cubes [16] or recover the entire signed distance field with fast-marching [1]. As previously mentioned, we want to avoid fast-marching to save time and memory. Therefore, the level set evolution scheme must restrict its computational domain to Λ_n for each level set φ_n. The sparse-field LSM [14] has exactly this property. Fig. 1 depicts both the unsigned distance field and label image representing 5 objects that overlap to create a total of $N = 8$ object regions.

| (a) | (b) | (c) | (d) |
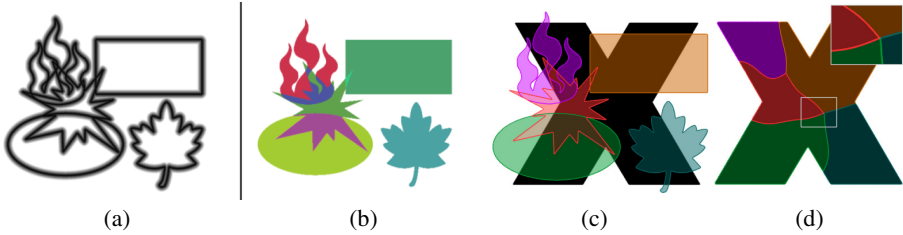
Fig. 1. (a) Distance field $\psi(x)$, (b) label image $\chi(x)$, (c) initial segmentation overlaid on "X" image, and (d) final segmentation of "X" image

Level Set Evolution. The N objects represented by N level sets $\vec{\varphi} = \{\varphi_1(x), \cdots, \varphi_N(x)\}$ are evolved through time by solving the differential equation $\frac{\partial \vec{\varphi}(x,t)}{\partial t} = \vec{f}(x,t)\boldsymbol{\delta}(\vec{\varphi}(x,t))$ where $\vec{f}(x,t) \in \mathcal{R}^N$ is the speed function and $\boldsymbol{\delta}(\cdot)$ is an $N \times N$ diagonal matrix whose diagonal entries are compactly supported approximations to the dirac delta $\delta(\cdot)$. The ith diagonal entry $\boldsymbol{\delta}_{ii}(\vec{\varphi}) = \delta(\varphi_i)$. Subsequent examples use speed functions of the following form:

$$f_n(x) = \lambda_\rho \rho(x)\|\nabla \varphi_n(x)\| + \lambda_v \vec{v}(x) \cdot \nabla \varphi_n(x) + \lambda_\kappa \kappa_n(x)\|\nabla \varphi_n(x)\|,$$

$$(3)$$

where $\rho(x) \in \mathcal{R}$ is a pressure force that drives the object's boundary towards a particular image intensity, $\vec{v}(x) \in \mathcal{R}^d$ is an external velocity field that drives the boundary towards edges in the image, and $\kappa_n(x) \in \mathcal{R}$ is the mean curvature for object n. Relative contributions of each force are controlled by weights λ_ρ, λ_v, and λ_κ. Forces are computed with first-order upwind finite differences [1] on either a 3×3 or $3 \times 3 \times 3$ stencil in 2D or 3D respectively. A first-order solution to $\frac{\partial \vec{\varphi}(x,t)}{\partial t}$ yields the following iterative scheme:

$$\vec{\varphi}(x, t + \Delta t) = \vec{\varphi}(x,t) + \Delta t \vec{f}(x,t)\boldsymbol{\delta}(\vec{\varphi}(x,t)).$$

$$(4)$$

The evolution process changes the location of the zero iso-level for each object, thereby moving object boundaries in accordance with the speed function $\vec{f}$. After computing updates for all level sets $\vec{\varphi}(x)$, the result must be stored in the label and unsigned distance images $\chi(x)$ and $\psi(x)$ respectively. To do this, we use the projection method

proposed by Losasso et al. (See Fig. 2,4,5 in [8]) . New level set values are sorted to find the smallest two $\boldsymbol{\varphi_a}(\boldsymbol{x})$ and $\boldsymbol{\varphi_b}(\boldsymbol{x})$ s.t. $\boldsymbol{\varphi_a}(\boldsymbol{x}) < \boldsymbol{\varphi_b}(\boldsymbol{x})$ among those labels $\boldsymbol{n}$ for which $\boldsymbol{x} \in \Lambda_{\boldsymbol{n}}$. The label and unsigned distance images are updated via:

$$\chi(\boldsymbol{x},t) = \begin{cases} b & \varphi_a(\boldsymbol{x},t) = \varphi_b(\boldsymbol{x},t) \text{ and } b < a \\ a & \text{otherwise} \end{cases} \tag{5}$$

and

$$\psi(\boldsymbol{x},t) = \tfrac{1}{2}|\varphi_a(\boldsymbol{x},t) - \varphi_b(\boldsymbol{x},t)|. \tag{6}$$

This projection technique has several useful properties. First, it reduces the complexity of topological relationships between N objects in any finite dimension to just two objects in one dimension. Second, it insures objects within a distance of 1 pixel from each other do not overlap or have air-gaps. Third, it couples forces between adjacent objects so that the shared boundary can move without creating gaps or overlaps.

The initial segmentation may have overlaps even though the final image segmentation should not. One goal of level set evolution is to remove these overlaps. For this task, overlapping object regions in the initial segmentation are treated as different objects. It is then necessary to define forces that contract overlapping object regions so that the final segmentation contains only distinct objects.

To evolve $\tilde{\varphi}_n(\boldsymbol{x})$, we evaluate eq. 3 on the subset $\Gamma_n = \{\boldsymbol{x} | -0.5 \leq \tilde{\varphi}_n(\boldsymbol{x}) \leq 0.5, \forall \boldsymbol{x} \in \Omega\} \subseteq \Lambda_n$, by windowing $\delta(\cdot)$ to have support $[-0.5, 0.5]$ pixels. The CFL condition is enforced by choosing Δt s.t. $\Delta t \leq 0.5/f_{max}$ where $f_{max} = \max_{n,x}|f_n(\boldsymbol{x})\delta(\psi(\boldsymbol{x},t))|$. To compute $\vec{f}(\boldsymbol{x},t)$ with finite differences, $\varphi_n(\boldsymbol{x})$ must be known in the neighborhood around $\boldsymbol{x}$. If $\boldsymbol{x} \in \Lambda_n$ and $\chi(\boldsymbol{x}) = n$, it is true that $\forall \boldsymbol{y} \in \mathcal{N}(\boldsymbol{x})$, $\varphi_n(\boldsymbol{y}) = \sigma_n(\boldsymbol{y})\psi(\boldsymbol{y})$ because either $\chi(\boldsymbol{y}) \neq \chi(\boldsymbol{x})$ which implies $\boldsymbol{y} \in \Lambda_n$, or $\chi(\boldsymbol{y}) = \chi(\boldsymbol{x})$ which implies $\boldsymbol{y}$ is inside object n, so $\psi(\boldsymbol{y})$ must be a measurement to object n. If $\boldsymbol{x} \in \Lambda_n$ but $\chi(\boldsymbol{x}) = m$ and $\chi(\boldsymbol{y}) = l$ s.t. $n, m,$ and l are all different, $\psi(\boldsymbol{y})$ could be a measurement to object l instead of n. The problem can be resolved by setting $\varphi_n(\boldsymbol{y}) = (\psi(\boldsymbol{x}) + 1)\sigma_n(\boldsymbol{y})$ if $\boldsymbol{y} \notin \Lambda_n$. To address the problem in general for larger neighborhoods, we recommend the Multi-compartment LSM [5, 9]. A solution is not implemented in this work because the problem is unnoticeable in practical examples. Future work will implement a solution.

The level set evolution process is described by *Algorithm 1*. The unsigned distance field ψ is rebuilt within a distance of $L = 3$ pixels via the fast approximation described in [14]. **Evolve** is straightforward to parallelize because each for-loop over Ω is dependent on only the connected neighborhood around each pixel. The only step that is non-trivial to parallelize is computation of f_{max}, which can be done with a parallel reduction. This computation can be avoided by crafting (rescaling) $\vec{f}$ s.t. $f_n(\boldsymbol{x},t) \in [-1,1]$ or clamping $\tau_k(\boldsymbol{x})$ to the range $[-1,1]$. Clamping is an acceptable shortcut in image segmentation problems because forces do not have to be physically accurate. Furthermore, forces only need to be evaluated for objects that compete for a particular pixel, which is at most $(2d + 1)$. The computational complexity of **Evolve** is then $O(M^d/P)$ for M^d pixels and P processing units. Details of the implementation are contained in the open-source release, which does compute f_{max}.

Algorithm 1. **Evolve**

foreach $x \in \Omega$ **do** //Compute speeds
    if $\exists n$ s.t. $x \in \Gamma_n$**then**
  foreach $y_k \epsilon (\mathcal{N}(x) \cup \{x\})$
        $n = \chi(y_k, t); \tau_k(x) = f_n(x, t) \delta(\varphi_n(x, t)); l_k(x) = n$
$f_{max} = \max_{k,x} |\tau_k(x)|$
if $f_{max} > 1$ **then** $\Delta t = 0.5/f_{max}$ **else** $\Delta t = 0.5$ //Determine time step
foreach $x \in \Omega$ **do** //Update level sets
    if $\exists n$ s.t. $x \in \Gamma_n$**then** $Z = \emptyset$**;**
        for $k = 1: (2d + 1)$ **do**
          $n = l_k(x)$**;**$z = \tilde{\varphi}_n(x, t) + \Delta t \tau_k(x)$
          $Z = Z \cup \{(z, n)\}$
        Sort Z **by** z **to find** $\varphi_a(x)$ **and** $\varphi_b(x)$
Compute $\chi(x, t + \Delta t)$ **and** $\psi(x, t + \Delta t)$
for $l = 1: L$ **do** //Rebuild distance field
  foreach $x \in \Omega$ **do**
    if $\psi(x, t) > 0.5$ **and** $\psi(x, t + \Delta t) > (l - 0.5)$ **then**
    $\psi(x, t + \Delta t) = \min_{y \epsilon \mathcal{N}(x)} |\sigma_{\chi(x)}(y, t + \Delta t) \psi(y, t + \Delta t) - 1|$

3 Results

Image Segmentation 2D. Multi-Object Geodesic Active Contours (MOGAC) were applied to segmentation of 2D images into multiple compartments. The following segmentation example of a 512×512 "X" image (Fig. 1c) was constructed to evaluate the performance of the algorithm. In the first experiment, the image was segmented into 5 objects (i.e. compartments). Objects overlap in the initial segmentation for a total of $N = 8$ labeled object regions. The 5 objects were evolved with pressure and curvature forces [2]. Overlapping object regions were assigned a constant inward pressure force to cause their contraction. The final segmentation contains only the 5 objects with no gaps or overlaps (Fig. 1d).

To evaluate scalability of the algorithm, the problem size was increased by horizontally and vertically tiling the "X" image and initial segmentations. The algorithm was executed on a PC with dual Intel E5630s and NVIDIA Quadro 4000. Fig. 2a shows the computation time per iteration averaged over the 2000 iterations required to segment each image. The original image took 4.9 sec. to segment on the GPU, and the computation time scaled almost linearly as a function of image size.

A second experiment was conducted to evaluate the algorithm's performance as a function of the number of objects. Segmentation of the original image was initialized with between 1 and 16 randomly placed circles. Results are depicted in Fig. 2b. Computation time is almost constant as a function of the number of objects.

MOGAC was applied to cell tracking in a 1024×1024 microscopy image acquired from the Cell Centered Database (CCDB) [17]. The image was first automatically segmented with Medical Image Processing And Visualization (MIPAV)
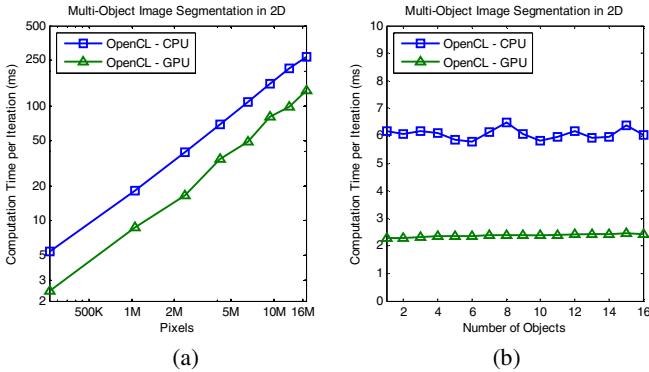
Fig. 2. Segmentation of "X" image as a function of (a) image size and (b) number of objects

[18] through gray-level morphology and thresholding techniques (Fig. 3a). The segmentation was then refined with MOGAC to better localize boundaries on the 138 detected objects (Fig. 3b). The segmentation was driven by pressure forces and external velocity field produced by Gradient Vector Flow (GVF) [19]. The GPU implementation of MOGAC ran for 250 iterations for 3.03 sec. total.
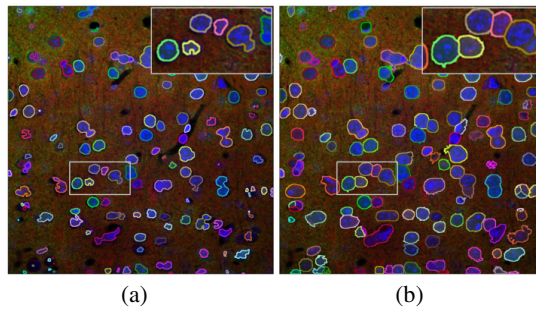
Fig. 3. (a) Initial segmentation of microscopy image and (b) MOGAC segmentation refinement of 138 objects. Cells appear blue in this imaging modality

Image Segmentation 3D. The 3D version of MOGAC is almost identical to the 2D version, except that a 6-connected 3D neighborhood is used and finite difference calculations are evaluated on a $3 \times 3 \times 3$ stencil. MOGAC was used to clean-up manual segmentations of MR images. In the first experiment, MOGAC was initialized with a manual segmentation of the epicardium in an $128 \times 256 \times 128$ MR image that contains small gaps and overlaps between structures (Fig. 4a,b). MOGAC was used to remove these gaps and overlaps to produce a sub-voxel segmentation that is a proper partition of the epicardium into 4 structures (Fig. 4c,d). The MOGAC clean-up required 50 iterations (2.5 sec. total). In the second experiment, a whole brain was segmented into 10 structures with TOADS [20] on a $256 \times 256 \times 256$ MR image from the OASIS database [21]. The hard classification (Fig. 5a) was smoothed with MOGAC for 10 iterations (2.3 sec. total) to produce the segmentation shown in Fig. 5b.

(a)   (b)   (c)   (d)
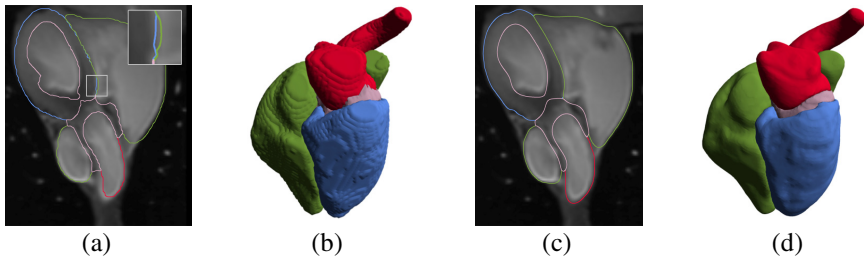
Fig. 4. (a,b) Initial heart segmentation and (c,d) MOGAC segmentation. Epicardium showing right ventricle and atrium (green), myocardium (blue), left ventricle (pink), and left atrium (red).

(a)   (b)
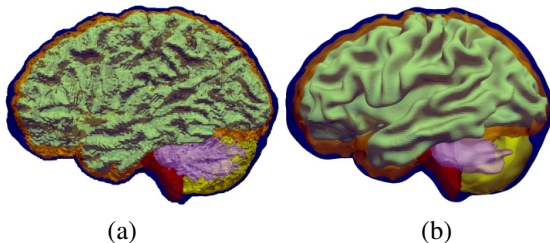
Fig. 5. (a) Initial hard segmentation (voxel precision) of 10 brain structures. (b) Segmentation after mean curvature smoothing with MOGAC (sub-voxel precision). It is common and beneficial to use voxel precision methods (a) for speed and follow them with sub-voxel precision methods (b) for smoothness. The Multi-compartment LSM [5, 9] is reported to require 22.5 sec per iteration to perform this same task. MOGAC is approximately 100x faster.

4   Discussion

A Multi-object LSM has been presented that can segment any number of objects with the same small memory footprint. The "label mask + distance field" data structure does not have to be "decompressed" into individual level sets either before, during, or after segmentation. Table 1 summarizes the theoretical performance for each Multi-object LSM. MOGAC is theoretically faster and uses less memory than existing methods. Because the MOGAC algorithm is structurally similar to single object LSM, GPU accelerated LSM techniques described in other works are applicable [15].

Algorithm 1 spends a lot of computation time checking if $\exists n$ s.t. $x \in \Gamma_n$. To avoid traversing the entire volume, a more work-efficient approach is to index Γ_n. A parallel algorithm for indexing Γ_n has already been described [15], and a variant of that algorithm is included in the public release (Work-efficient MOGAC).

Table 1. Algorithm complexity for Multi-object LSMs based on M^d pixels, N objects, and P processing units in dimension $d = \{2,3\}$

Methods	Time	Memory
N level set methods [4, 6, 8]	$O(NM^d \log M)$	$O(NM^d)$
Multi-phase [7]	$O(M^d \log M \log N)$	$O(M^d \log N)$
Multi-compartment [5, 9]	$O(M^d \log(M/N))$	$O(M^d)$
MOGAC	$O(M^d/P)$	$O(M^d)$
Work-efficient MOGAC	$O(M^{d-1}/P)$	$O(M^d)$

References

1. Sethian, J.: Level set methods and fast marching methods. Cambridge Univ. Pr. (1999)
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22, 61–79 (1997)
3. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. Int. J. of Computer Vision 72, 195–215 (2007)
4. Samson, C., Blanc-Féraud, L., Aubert, G., Zerubia, J.: A level set model for image classification. Int. J. of Computer Vision 40, 187–197 (2000)
5. Fan, X., Bazin, P.L., Prince, J.L.: A multi-compartment segmentation framework with homeomorphic level sets. Presented at the IEEE Conf. on Computer Vision and Pattern Recognition (2008)
6. Paragios, N., Deriche, R.: Coupled Geodesic Active Regions for Image Segmentation: A Level Set Approach. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1843, pp. 224–240. Springer, Heidelberg (2000)
7. Vese, L.A., Chan, T.F.: A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. Journal of Computer Vision 50, 271–293 (2002)
8. Losasso, F., Shinar, T., Selle, A., Fedkiw, R.: Multiple interacting liquids. ACM Transactions on Graphics (TOG) 25, 812–819 (2006)
9. Fan, X., Bazin, P.L., Bogovic, J., Bai, Y., Prince, J.L.: A multiple geometric deformable model framework for homeomorphic 3D medical image segmentation. Presented at the Computer Vision and Pattern Recognition Workshops, CVPRW 2008 (2008)
10. Lie, J., Lysaker, M., Tai, X.C.: A binary level set model and some applications to Mumford-Shah image segmentation. IEEE Transactions on Image Processing 15, 1171–1181 (2006)
11. Adams, R., Bischof, L.: Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 641–647 (1994)
12. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 22, 888–905 (2000)
13. Li, K., Miller, E.D., Chen, M., Kanade, T., Weiss, L.E., Campbell, P.G.: Cell population tracking and lineage construction with spatiotemporal context. Medical Image Analysis 12, 546–566 (2008)
14. Whitaker, R.: A level-set approach to 3D reconstruction from range data. International Journal of Computer Vision 29, 231 (1998)
15. Roberts, M., Packer, J., Sousa, M.C., Mitchell, J.R.: A work-efficient GPU algorithm for level set segmentation. In: HPG 2010, pp. 123–132 (2010)
16. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. ACM Siggraph Computer Graphics 21, 163–169 (1987)
17. Martone, M.E., Gupta, A., Wong, M., Qian, X., Sosinsky, G., Ludäscher, B., Ellisman, M.H.: A cell-centered database for electron tomographic data. Journal of Structural Biology 138, 145–155 (2002)

18. McAuliffe, M.J., Lalonde, F.M., McGarry, D., Gandler, W., Csaky, K., Trus, B.L.: Medical image processing, analysis & visualization in clinical research. In: CBMS, p. 0381 (2001)
19. Chenyang, X., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7, 359–369 (1998)
20. Bazin, P.L., Pham, D.L.: Topology-preserving tissue classification of magnetic resonance brain images. IEEE Transactions on Medical Imaging 26, 487–496 (2007)
21. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open Access Series of Imaging Studies (OASIS). Journal of Cognitive Neuroscience 19, 1498–1507 (2007)

A Pattern Recognition Approach to Zonal Segmentation of the Prostate on MRI

Geert Litjens, Oscar Debats, Wendy van de Ven,
Nico Karssemeijer, and Henkjan Huisman

Radboud University Nijmegen Medical Centre
Geert Grooteplein-Zuid 10, 6525GA Nijmegen, The Netherlands

Abstract. Zonal segmentation of the prostate into the central gland and peripheral zone is a useful tool in computer-aided detection of prostate cancer, because occurrence and characteristics of cancer in both zones differ substantially. In this paper we present a pattern recognition approach to segment the prostate zones. It incorporates three types of features that can differentiate between the two zones: anatomical, intensity and texture. It is evaluated against a multi-parametric multi-atlas based method using 48 multi-parametric MRI studies. Three observers are used to assess inter-observer variability and we compare our results against the state of the art from literature. Results show a mean Dice coefficient of 0.89 ± 0.03 for the central gland and 0.75 ± 0.07 for the peripheral zone, compared to 0.87 ± 0.04 and 0.76 ± 0.06 in literature. Summarizing, a pattern recognition approach incorporating anatomy, intensity and texture has been shown to give good results in zonal segmentation of the prostate.

Keywords: prostate, MRI, segmentation, voxel classification, atlas.

1 Introduction

Prostate cancer is a major health problem in the Western world, with one in six men affected during their lifetime [1]. Multi-parametric magnetic resonance imaging (MPMR) has been shown to play an important role in the diagnosis of prostate cancer [2]. A typical MR exam contains T2-weighted, dynamic-contrast-enhanced and diffusion-weighted imaging. Interpretation of MPMR prostate studies is challenging, and therefore the use of computer-aided diagnosis techniques has been investigated [3]. For correct interpretation of MPMR knowledge about the zonal anatomy of the prostate is required, because the occurrence and appearance of cancer is dependant on its zonal location [4]. From a radiological point of view the prostate is usually considered to have two visible zones on MRI, the central gland (CG) and the peripheral zone (PZ) [5]. We are exploring options to integrate knowledge about the zonal anatomy into CAD systems. For this automated segmentation of the zones is the first step. The availability of zonal segmentation is also mandatory for those CAD methods in literature that focus on the PZ only, as for example in [3].

Although much research has been done on prostate segmentation [6,7], only recently the first study on segmentation of the individual zones was published by Makni et al. [8]. In their study they investigated the use of an evidential C-means clustering (ECM) approach to cluster voxels into their respective zones. In addition, they extended the ECM approach to incorporate the spatial relation between voxels. Using this method they obtained good results on their data set (0.87 ± 0.04 mean Dice coefficient for the central gland compared to a simultaneous truth and performance level estimation (STAPLE) obtained ground truth [9]). To the best of the authors knowledge their paper remains the only published paper evaluating prostate zonal segmentation.

The purpose of this paper was to investigate a pattern recognition algorithm to segment the prostate zones. The pattern recognition approach uses several image features with a voxel classifier to detect the zones. This is a method that has been explored in many other segmentation problems. We compare it to a multi-parametric multi-atlas approach which is used to simultaneously segment the prostate and the prostate zones. Additionally, we will compare our results to inter-observer variability and the results obtained by Makni et al.[8]

2   Methods

2.1   Multi-parametric Multi-atlas Segmentation

Multi-atlas segmentation is an accurate method for prostate segmentation, as has been shown by Klein et al. [6] We have chosen a similar approach, but extended it to use multi-parametric data. We evaluated the atlas method with both majority voting and STAPLE [9] to obtain the final binary segmentation.

The registration of the atlases to the new case is performed using the elastix software package [10]. For the registration we use local normalized mutual information as a similarity metric. We register both the T2-weighted image and the quantitative apparent diffusion coefficient (ADC) map simultaneously. We chose to add the ADC map to the registration because it contains additional information on the zonal distribution within the prostate. In a previous experiment we investigated the added value of the ADC in zonal segmentation and we noticed that it improved performance. The cost function we then optimize can be expressed as

$$C(T_\mu; I_F, I_M) = \frac{1}{\sum_{i=1}^{N} \omega_i} \sum_{i=1}^{N} \omega_i C(T_\mu; I_F^i, I_M^i) \tag{1}$$

were C is the cost function, T_μ is the registration transformation, I_F is the fixed (the unknown case) and I_M the moving image (the atlas). Furthermore, ω_i is the weight for each of the multi-parametric images i were $i = 1$ is the T2-weighted image and and $i = 2$ is the ADC map. We chose ω to be 0.5 for both i.

The registration consists of two distinct steps. In the first step we register using only a translation transform to align the images to the new case. The second step is an elastic registration using a b-spline transformation. After the

(a) T2-Weighted image
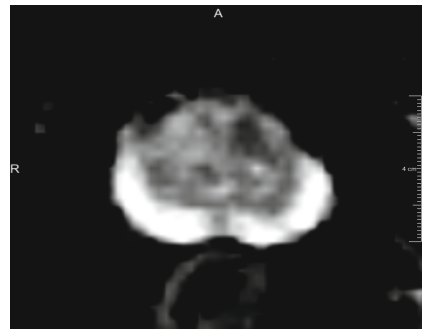
(b) Apparent diffusion coefficient map
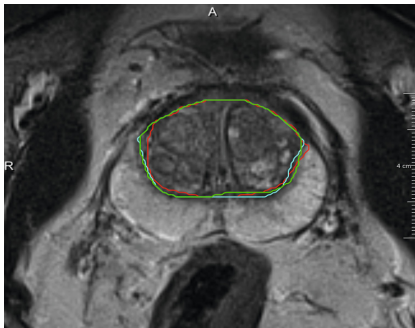
(c) Central gland observer segmentation (Red, cyan and green for observer 1,2 and 3)
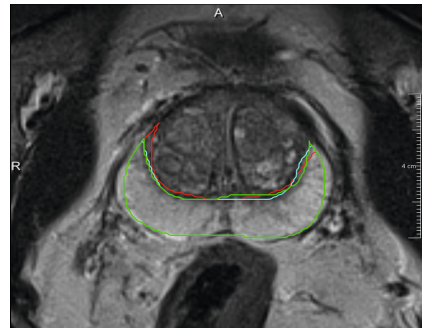
(d) Peripheral zone observer segmentation (Red, cyan and green for observer 1,2 and 3)
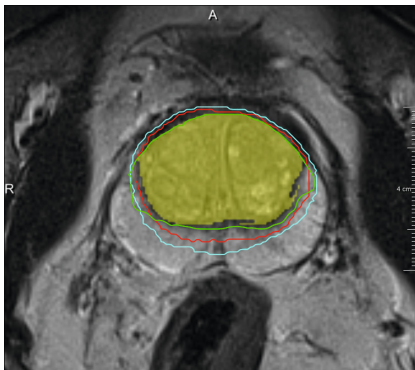
(e) Central gland automatic segmentation (Red, cyan and green for atlas (voting), atlas (STAPLE) and voxel classification), the STAPLE constructed 'true' segmentation is overlayed in yellow
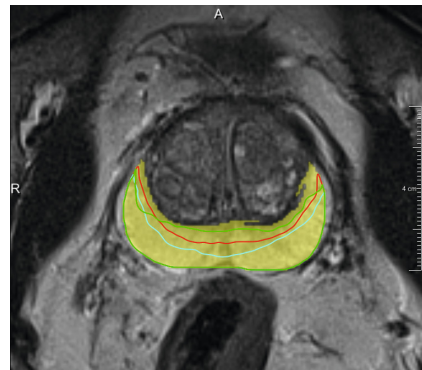
(f) Peripheral zone automatic segmentation (Red, cyan and green for atlas (voting), atlas (STAPLE) and voxel classification), the STAPLE constructed 'true' segmentation is overlayed in yellow

Fig. 1. Example data set with T2-W image and ADC map in a and b and segmentation results in c, d, e and f

registration the obtained transformation is used to transform the known binary segmentations to the target image space. These can subsequently be used to construct the unknown binary segmentation. Several approaches exist in literature, of which majority voting is the simplest and best known method [6]. We compare this approach with optimizing the segmentation by using STAPLE [9].

2.2   Voxel Classification Segmentation

For the voxel classification segmentation we determined a set of features that represent the difference between the two zones. These features can be separated into three categories: anatomy (positional), intensity and texture.

For the anatomy features we use the information we know from the normal prostate composition. The peripheral zone is usually situated at the dorsal side of the prostate, getting thicker towards the apex of the prostate. We chose to model this by developing a set of three relative position and distance features. Given the whole prostate mask we can calculate a relative position in each direction for each voxel, resulting in a value between 0 and 1. We calculate this feature in the ventrodorsal direction and the craniocaudal direction. In addition, the relative distance (also between 0 and 1) to the prostate boundary is given as a feature.

Two intensity features are included in the voxel classification step. The first intensity feature we use is the apparent diffusion coefficient (ADC) for each voxel, which itself should be a quantitative feature. The second intensity feature we use is a calculated T2 value for each voxel. Using the T2 relaxation time instead of the T2-weighted voxel values will make this feature much more robust to changes in scan parameters. To this end we used the following signal model equation for turbo-spin-echo sequences:

$$\text{T2}_{\text{p}} = -\text{TE} \left(\log e^{\frac{-\text{TE}}{\text{T2}_{\text{m}}}} \frac{S_{\text{m}}^{\text{PD}} S_{\text{p}}^{\text{T2W}}}{S_{\text{m}}^{\text{T2W}} S_{\text{p}}^{\text{PD}}} \right)^{-1} \tag{2}$$

Here T2 is the estimated T2 relaxation time, TE is the echo time for the MR pulse sequence, S the signal intensity. The superscript PD and T2W represent either the proton density weighted image or the T2-weighted image. The subscript p and m denote prostate and muscle respectively. Using this equation and a region of interest placed in a skeletal muscle we can calculate the true T2 relaxation time for each voxel given the proton density and T2-weighted images.

The muscle ROI is automatically selected using a search method. Starting from the bottom slice of the T2-weighted image an Otsu threshold is performed to separate the dark areas (including the muscles) from the bright areas. We are looking for the two muscles alongside the prostate, so we suppress the center of the image with a rectangular block. Then a connected component analysis is used to find individual dark components in the image. The two largest connected components should correspond to the left and right muscle. We make sure this is the case by investigating the shape and symmetry of the two connected components. The muscle are less wide than long and they should have approximately the same shape on the left and right. We mirror the left connected component

and investigate the Jaccard index with the right connected component. The minimum value for width divided by the length is 0.75 and the threshold for the Jaccard index is 0.5. The resulting connected components are eroded to ensure that the ROI is completely in the muscle.

The third set of features consists of five texture features. The first two features are homogeneity and correlation calculated using the co-occurrence matrix [11]. We used 16 gray value bins for the histogram and took the average over all 2D directions. The third and fourth feature are entropy and texture strength, based on the Neighborhood Gray-Tone Difference Matrix [12]. Here also 16 gray level bins were used, in combination with an evaluation distance of 1. For all of these features the kernel size was 10x10x1 voxels. The fifth feature was the local binary pattern at each voxel [13], which was calculated over a 3x3x1 voxel neighborhood. For this feature the images were down-sampled using Gaussian re-sampling such that a 3x3x1 neighborhood corresponded to a 12x12x1 neighborhood.

After calculating the features a balanced training set is constructed. Hard classification using a linear discriminant classifier is performed to obtain a binary segmentation of the central gland. To smoothen the initial boundary some post-processing is performed. Firstly, connected component analysis is used to select the largest connected component. Erosion and dilation are then performed to remove small objects attached to the segmentation. Finally the edge voxels between the central gland and the peripheral zone are selected and a thin plate spline is fitted through these voxels. This results in our final segmentation.

3   Validation

For validation we used 48 multi-parametric MR studies with manual segmentations of the whole prostate. For each case the transversal T2-weighted scan (resolution 0.6x0.6x4 mm) and the apparent diffusion coefficient map (2x2x4 mm) were used. In addition, for the voxel classification step, the proton density weighted image was used to calculate the T2 values. The ADC and proton density images were inspected to assess the alignment with the T2-weighted image. If needed, they were corrected to obtain good alignment.

The ground truth was constructed by STAPLE [9] to merge the manual segmentations done by three observers. The observers made manual segmentations by indicating the zonal boundary on each T2-weighted image slice given the manual whole prostate segmentation. We validated the automatic segmentations by calculating three similarity measures: the Jaccard index (JI), the Dice similarity coefficient (DSC) and the volume difference (VD). The Jaccard index is given as $J = \frac{|V_1 \cap V_2|}{|V_1 \cup V_2|}$, were V_1 and V_2 are the automated segmentation and the STAPLE ground truth respectively. The Dice coefficient is similar to the Jaccard index and can be expressed as $D = \frac{2|V_1 \cap V_2|}{|V_1| + |V_2|}$. Lastly, the volume difference can be expressed as VD $= |V_1| - |V_2|$. Validation was performed in a leave-one-out manner, thus the case to be segmented was removed from the set of atlases for the atlas method and from the training data for the voxel classification.

4   Results

In figures 2a, 2c and 2e the results of the segmentations of the central gland are presented. An example case is also shown in figure 1. We can see that the observers all perform well with respect to the STAPLE ground truth. For the segmentation methods the voxel classification approach outperforms the atlas based methods (mean DSC 0.89 ± 0.03 vs 0.80 ± 0.013 for majority voting and 0.80 ± 0.17 for STAPLE), although it is not as good as the human observers (mean DSC's 0.95 ± 0.06, 0.97 ± 0.05, 0.96 ± 0.06). The JI and VD (figure 2b and figure 2c) show similar results. The VD results show that our methods in general under-segment the central gland. If we compare our results to those in Makni et al. [8] we perform slightly better using our voxel classification approach, as they report a mean DSC of 0.87 ± 0.04. For the peripheral zone we see similar results (figures 2b, 2d and 2f). Our pattern recognition approach outperforms the atlas based method and is relatively close to the observer scores. Here the pattern recognition approach has a mean DSC of 0.75 ± 0.07 compared to 0.82 ± 0.15, 0.89 ± 0.12 and 0.86 ± 0.11 for the observers. The atlas methods both perform poorly with respect to the peripheral zone with a mean DSC of 0.57 ± 0.19 and 0.48 ± 0.22. Compared to the state of the art we perform slightly worse, with a mean DSC of 0.76 ± 0.06 compared to our 0.75 ± 0.07.

5   Discussion

In this paper we investigated a pattern recognition approach to zonal segmentation of the prostate. We compared our method to an atlas based method and to the method published by Makni et al. Our results show that the voxel classification method outperforms the atlas based method. It also shows similar performance compared to the method published by Makni et al. We believe the pattern recognition approach outperforms the atlas-based method because it is less restrictive than an atlas, which is limited to the shapes available within the atlases. Additionally, pattern recognition allow for non-linear combination of all features, including texture features.

This study also has limitations. A true comparison with the results from Makni et al. is difficult, mostly due to differences in the data used, for example in resolution. Additionally, for the atlas method we did not use the manual whole prostate segmentations because this method segments the whole prostate and the zones at the same time. This might cause some bias compared to the voxel classification approach were we did use the whole prostate manual segmentation. We did investigate using the manual whole prostate mask for the atlas method by only evaluating the registration metric within the mask. However, this approach gave worse results than not using the whole prostate mask at all. Both methods performed worst when the peripheral zone is very thin, then partial volume effects and unclear boundaries between the zones make it difficult to segment them. Finally, our voxel classification approach might be improved by incorporating additional texture features (e.g. Gaussian or Gabor based texture features) or by incorporating global information like prostate volume [8].
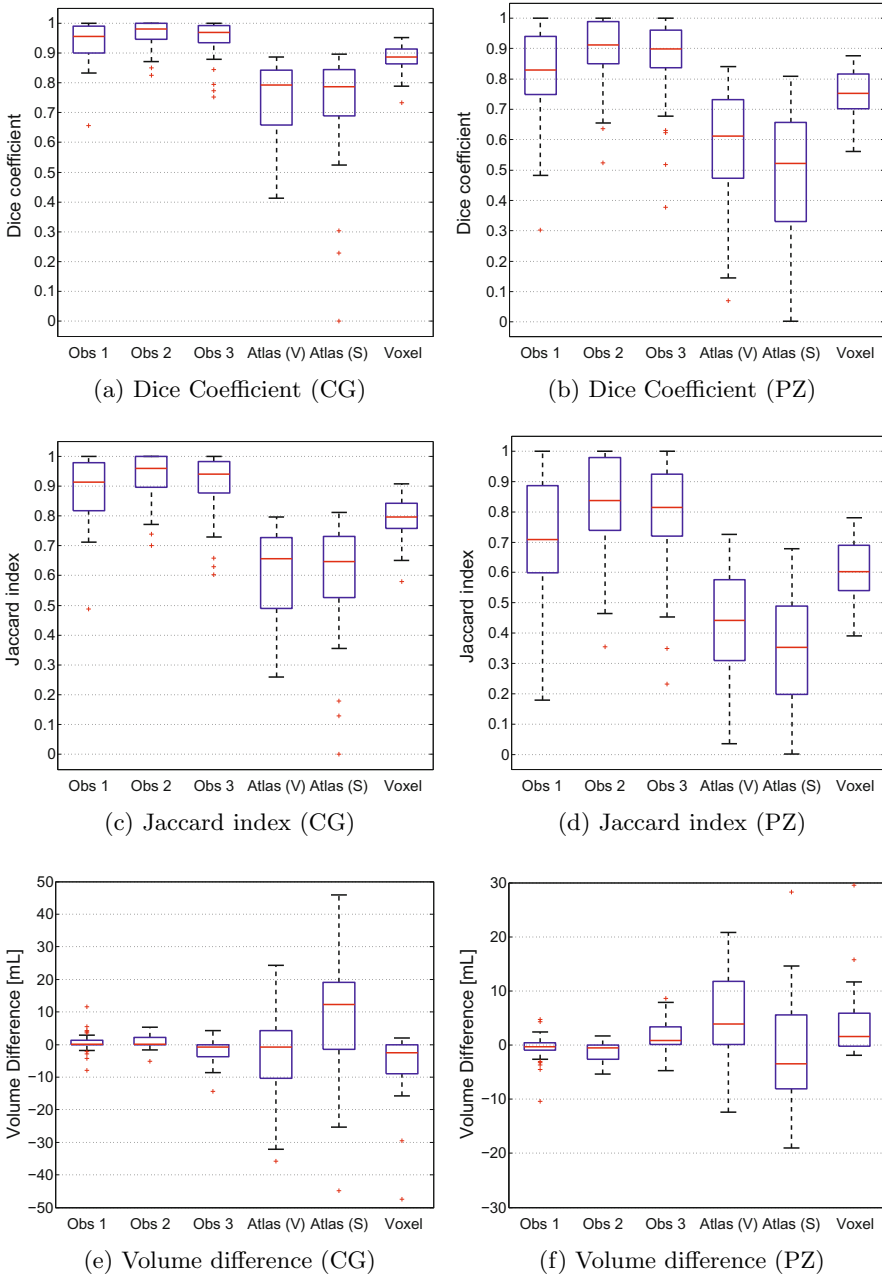
Fig. 2. Results of the segmentation methods. The captions on the x-axes correspond to observers 1, 2 and 3, the atlas method using majority voting, the atlas method using STAPLE and the voxel classification approach.

Summarizing, a new pattern recognition approach to segment the prostate zones was presented, incorporating anatomical, intensity and texture features. It outperforms an atlas based method, is relatively close to the inter-observer performance and shows similar performance compared to the state of the art.

References

1. Siegel, R., Naishadham, D., Jemal, A.: Cancer statistics, 2012. CA Cancer J. Clin. 62, 10–29 (2012)
2. Kitajima, K., Kaji, Y., Fukabori, Y., Yoshida, K., Suganuma, N., Sugimura, K.: Prostate cancer detection with 3 T MRI: comparison of diffusion-weighted imaging and dynamic contrast-enhanced MRI in combination with T2-weighted imaging. J. Magn. Reson. Imaging 31, 625–631 (2010)
3. Chan, I., Wells, W., Mulkern, R.V., Haker, S., Zhang, J., Zou, K.H., Maier, S.E., Tempany, C.M.C.: Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. Med. Phys. 30, 2390–2398 (2003)
4. Viswanath, S.E., Bloch, N.B., Chappelow, J.C., Toth, R., Rofsky, N.M., Genega, E.M., Lenkinski, R.E., Madabhushi, A.: Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo T2-weighted MR imagery. J. Magn. Reson. Imaging (February 2012)
5. Villeirs, G.M., Verstraete, K.L., De Neve, W.J., De Meerleer, G.O.: Magnetic resonance imaging anatomy of the prostate and periprostatic area: a guide for radiotherapists. Radiother. Oncol. 76(1), 99–106 (2005)
6. Klein, S., van der Heide, U.A., Lips, I.M., van Vulpen, M., Staring, M., Pluim, J.P.W.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med. Phys. 35, 1407–1417 (2008)
7. Toth, R., Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., Kalyanpur, A., Pungavkar, S., Madabhushi, A.: A magnetic resonance spectroscopy driven initialization scheme for active shape model based prostate segmentation. Med. Image Anal. 15, 214–225 (2011)
8. Makni, N., Iancu, A., Colot, O., Puech, P., Mordon, S., Betrouni, N.: Zonal segmentation of prostate using multispectral magnetic resonance images. Med. Phys. 38, 6093 (2011)
9. Warfield, S.K., Zou, K.H., Wells, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans. Med. Imaging 23, 903–921 (2004)
10. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.W.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205 (2010)
11. Amadasun, M., King, R.: Textural features corresponding to textural properties. IEEE Trans. Syst. Man. Cybern. 19, 1264–1274 (1989)
12. Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., Chinander, M.R.: Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. Acad. Radiol. 12, 863–873 (2005)
13. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24, 971–987 (2002)

Statistical Shape Model Segmentation and Frequency Mapping of Cochlear Implant Stimulation Targets in CT

Jack H. Noble[1], René H. Gifford[2], Robert F. Labadie[3], and Benoît M. Dawant[1]

[1] Depts. of Elect. Eng. and Comp. Sci.,
[2] Hearing and Speech Sci.,
[3] Otolaryngology – Head & Neck Surg.,
Vanderbilt University, Nashville, TN 37235, USA
{jack.h.noble,rene.h.gifford,robert.labadie,
benoit.dawant}@vanderbilt.edu

Abstract. Cochlear implant (CI) surgery is considered standard of care treatment for severe hearing loss. CIs are currently programmed using a one-size-fits-all type approach. Individualized, position-based CI programming schemes have the potential to significantly improve hearing outcomes. This has not been possible because the position of stimulation targets is unknown due to their small size and lack of contrast in CT. In this work, we present a segmentation approach that relies on a weighted active shape model created using microCT scans of the cochlea acquired ex-vivo in which stimulation targets are visible. The model is fitted to the partial information available in the conventional CTs and used to estimate the position of structures not visible in these images. Quantitative evaluation of our method results in Dice scores averaging 0.77 and average surface errors of 0.15 mm. These results suggest that our approach can be used for position-dependent image-guided CI programming methods.

Keywords: Spiral ganglion, modiolus, cochlear implant, weighted active shape model segmentation.

1 Introduction

Cochlear Implants (CIs) are considered standard of care treatment for severe-to-profound sensory-based hearing loss. CIs restore hearing by applying electric potential to neural stimulation sites in the cochlea with an implanted electrode array. An audiologist programs the CI by determining how it maps the spectrum of detected sound frequencies to the set of electrodes for each recipient. Programming is performed based solely upon patient response to judgments of perceived loudness for individual electrodes. This process can be difficult and time intensive, and the majority of potentially adjustable parameters are typically left at the default settings determined by the manufacturer. Because of this one-size-fits-all approach, standard programming methods may not result in optimal hearing restoration for all recipients.

In the natural hearing process, ear anatomy mechanically translates sound into vibrations of the basilar membrane, which separates the two principal cavities of the cochlea, the scala tympani and the scala vestibuli (see Figure 1). These vibrations
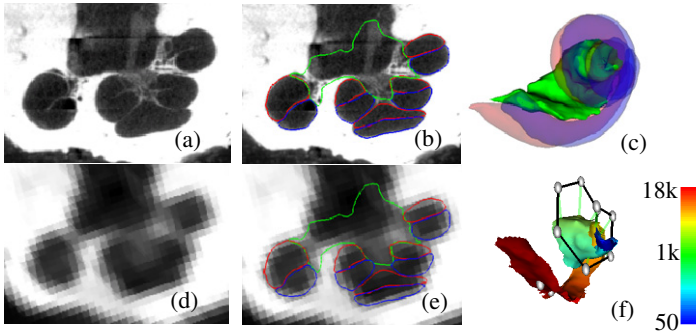
Fig. 1. Shown in (a) and (d) are a slice of a μCT and a CT of a human cochlea. In (b) and (e), the scala tympani (red), scala vestibuli (blue), and bundle of nerve cells of the SG (green) are delineated in the same slice. These structures are shown similarly in 3D in (c). In (f), the active region of the SG is colormapped with its tonotopy (Hz), and the positions of implanted electrodes are shown (gray spheres).

stimulate nerve cells connected to the spiral ganglion (SG) and, eventually, the auditory nerve. Researchers have extracted the tonotopic mapping between the frequency of the sound and the SG nerve cells that are stimulated, i.e., higher frequencies lead to stimulation of more basal SG nerve cells, whereas, lower frequencies stimulate more apical SG nerve cells [1]. CIs exploit this natural tonotopy by applying an electric field to more apical (basal) SG nerve cells to induce perceived lower (higher) frequency sounds. It is widely believed that programming schemes that use more explicit knowledge of electrode position hold great promise for improving hearing outcomes and could potentially be transformative to the field of CI audiology.

Since methods for identifying the position of implanted electrodes have already been established [2], the major obstacle for position-dependent programming is that there have been no techniques for accurately identifying the stimulation targets – the SG nerve cells. Identifying the SG in vivo is difficult because the nerve bundles have diameter on the order of microns and are too small to be visible in CT, which is the preferred modality for cochlear imaging due to its otherwise superior resolution (see Figure 1). However, the external walls of the cochlea are well contrasted in CT. Since the cochlea wraps around the SG, and, as shown in [3], external cochlear anatomy can be used to estimate the location of intra-cochlear anatomy using a statistical shape model (SSM); it is possible that a similar external anatomy driven SSM can be used to estimate the SG. In this work, we test such an approach for automatic segmentation and frequency mapping of the SG for computer assisted CI programming. The rest of this paper presents our approach.

2   Methods

The data set we have used in this study consists of images of six cadaveric cochlea specimens. For each specimen, we have acquired one μCT image volume with a SCANCO μCT. The voxel dimensions in these images are 36 μm isotropic. We also

have acquired CTs of the specimens with a Xoran xCAT® fpVCT scanner. In these volumes, voxels are 0.3 mm isotropic. In each of the μCT volumes, the scala vestibuli, scala tympani, and SG were manually segmented. Figure 1 shows an example of a conventional CT image and its corresponding μCT image.

2.1   Overview

Since the SG lacks any contrast in CT, we cannot segment it directly. The goal with our approach is to use the location of external cochlear features to predict the position of the SG. To do this, we have constructed a SSM of cochlear anatomy that includes the SG. Prior to constructing the SSM, we identify which points in the model will correspond to strong cochlear edges in CT. To those points we arbitrarily assign an importance weighting of 1. To all others we assign a lesser weighting of 0.01. These weights are used to construct a point distribution model (PDM) for weighted active shape model (wASM) segmentation [4]. The SSM is built as a standard PDM computed on the registered exemplar point sets. To segment a new image, the SSM is iteratively fitted in a weighted-least-squares sense to features in the target image. The edge points with their weight of 1 are fitted to strong edges in the CT. The non-edge points with low weight are fitted to the positions determined by non-rigid registration with an atlas image. With the weights that we have chosen, the non-edge points provide enough weak influence on the optimization to ensure that the wASM stays near the atlas-based initialization position, while the edge points strongly influence the whole wASM towards a local image gradient-based optimum for a highly accurate result. During model construction, the set of SG points in the model that interface with intra-cochlear anatomy were also identified. These points are referred to as the active region (AR) since they correspond to the region most likely to be stimulated by an implanted electrode. The tonotopic mapping of each point in the AR in the reference volume is computed based on angular depth using known equations [1]. Once a segmentation is completed, the tonotopic frequency labels from the model can be transferred to the target image. The following sub-sections detail these methods.

2.2   Model Creation

To model cochlear structures, we: (1) establish a point correspondence between the structures' surfaces, (2) use these points to register the surfaces to each other with a 7 degrees of freedom similarity transformation (rigid plus isotropic scaling), and (3) compute the eigenvectors of the registered points' covariance matrix. Point correspondence is determined using the approach described in [3]. Briefly, non-rigid registration is used to map each of the training volumes to a reference volume, and any errors seen in the results are manually corrected. Then, a correspondence is established between each point on the reference surface with the closest point in each of the registered training surfaces. Once correspondence is established, each of the training surfaces is point registered to the reference surface. Since the cochlear edge points will be the highest weighted points for the wASM segmentation, identical weights are used to register the training shapes in a weighted-least-squares sense using standard

point registration techniques [5] prior to computation of the eigenspace so that the model will best capture the shape variations at these points.

To build the model, the principal modes of shape variation are extracted from the registered training shapes. This is computed according to the procedure described by Cootes et. al. [6]: First, the covariance matrix of the point sets' deviation from the mean shape is computed as

$$C = \frac{1}{N}\sum_{j=1}^{N}(\vec{v}_j - \bar{v})(\vec{v}_j - \bar{v})^{\mathrm{T}}, \tag{1}$$

where the $\vec{v}_j$'s are the individual shape vectors and $\bar{v}$ is the mean shape. The shape vectors are constructed by stacking the 3D coordinates of all the points composing each structure into a vector. The modes of variation in the training set are then computed as the eigenvectors $\{\vec{u}_j\}$ of the covariance matrix,

$$\{\lambda_j, \vec{u}_j\}_{j=1}^{N-1} \quad : \quad \lambda_j\vec{u}_j = C\vec{u}_j. \tag{2}$$

These modes of variation are extracted for the combined shape of the scala tympani, scala vestibuli, and SG for all the samples in the training set.

2.3   Weighted Active Shape Segmentation

The procedure we use for segmentation with a wASM follows the traditional approach, i.e., (1) the model is placed in the image to initialize the segmentation; (2) better solutions are found while deforming the shape using weighted-least-squares fitting; and (3) eventually, after iterative shape adjustments, the shape converges, and the segmentation is complete. Initialization is performed using the atlas-based methods proposed in [3].

Once initialized, the optimal solution is found using an iterative searching procedure. At each search iteration, an adjustment is found for each model point, and the model is fitted in a weighted-least-squares sense, as described below, to this set of candidate adjustment points. To find the candidate points, two approaches are used. For cochlear edge points, candidates are found using line searches to locate strong edges. At each external point $\vec{y}_i$, a search is performed along the vector normal to the surface at that point. The new candidate point is chosen to be the point with the largest intensity gradient over the range of -1 to 1 mm from $\vec{y}_i$ along this vector. For all other points, it is impossible to determine the best adjustment using local image features alone because there are no contrasting features at these points in CT. Therefore, the original initialization positions for these points, which were provided by atlas-based methods, are used as the candidate positions. The atlas-based result, as our results will show, is sufficiently accurate to provide this useful information to the segmentation process.

The next step is to fit the shape model to the candidate points. We do this in the conventional wASM manner. A standard 7 degree of freedom weighted point registration is performed, creating similarity transformation T, between the set of candidate

points $\{\vec{y}_i{}'\}$ and the mean shape $\{\bar{v}_i\}$, where $\bar{v}_i$ are the 3D coordinates of the ith point in the mean shape. Then, the residuals

$$\vec{d}_i = T(\vec{y}_i{}') - \bar{v}_i \tag{3}$$

are computed. To obtain the weighted-least-squares fit coordinates in the SSM's eigenspace, we compute,

$$\vec{b} = (U^T W^T W U)^{-1} U^T W^T W \vec{d}, \tag{4}$$

where $\vec{d}$ is composed of $\{\vec{d}_i\}$ stacked into a single vector, $U = [\vec{u}_1 \quad \vec{u}_2 \quad ... \quad \vec{u}_{N-1}]$ is the matrix of eigenvectors, and W is a diagonal matrix with the importance point weightings in the appropriate entries along the diagonal. This equation results in a vector $\vec{b}$ that represents the coordinates in the SSM space corresponding to a weighted-least-squares fit of the model to the candidate points. The final approximation to the shape is computed by passing the sum of the scaled eigenvectors plus the mean shape through the inverse transformation, equivalently,

$$\vec{y}_i = T^{-1}\big(\bar{w}_i + \textstyle\sum_{j=1}^{N-1} b_j \vec{u}_{j,i}\big), \tag{5}$$

where $\vec{u}_{j,i}$ is the ith 3D coordinate of the jth eigenvector. As suggested by Cootes, the magnitude of the b_j's are constrained such that

$$\sqrt{\textstyle\sum_{j=1}^{N-1} \frac{b_j^2}{\lambda_j}} \leq 3, \tag{6}$$

which enforces the Mahalanobis distance between the fitted shape and the mean shape to be no greater than 3.

At each iteration, new candidate positions are found and the model is re-fitted to those candidates. The wASM converges when re-fitting the model results in no change to the surface. The tonotopic mapping of the SG points in the model, computed when the model was built, are directly transferred to the target image via the corresponding points in the converged solution. A result of this is shown in Figure 1f.

Validation. Segmentation was performed on CT's of the cochlea specimens using a leave-one-out approach, i.e., the volume being segmented is left out of the model. For one of the six specimens a CT was not available, and it was used as the model reference volume to simplify the leave-one-out validation study. Thus, the validation study measures segmentation error on the remaining five specimens when using PDMs with four modes of variation. Because these samples were excised specimens, rather than whole heads, the atlas-based initialization process required manual intervention—however, in practice the approach is fully automatic. To validate the results, we again exploit the set of corresponding μCT volumes. Each CT was rigidly registered to the corresponding μCT of the same specimen. The automatic segmentations were then projected from the CT to μCT space. Finally, Dice index of volume overlap [7] and
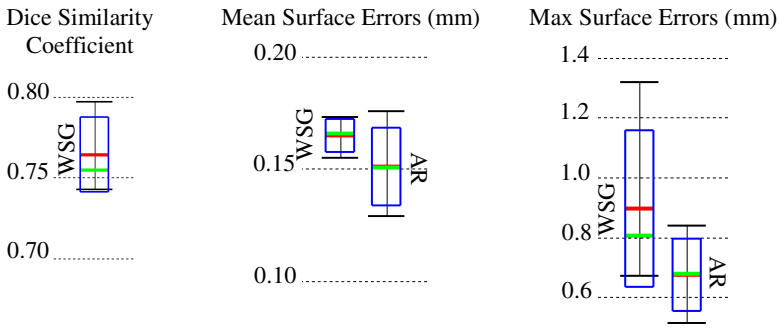
Fig. 2. Segmentation error distributions of Dice similarity scores for the whole SG (WSG) and mean and max symmetric surface error distributions for the WSG and in the active region (AR)

surface errors were computed between the registered automatic segmentations and the manual segmentations to validate the accuracy of our results. The distributions of surface error we report include distances in the forward (automatic-to-manual) direction as well as the reverse (manual-to-automatic) direction. A smaller distance between the automatic and manually segmented surfaces indicates higher accuracy.

3   Results

Quantitative comparisons between manual and automatic segmentations of the SG are shown in Figure 2. The Dice index and bi-directional mean/max surface distances were computed between each pair of automatic and manual segmentations. Figure 2 shows the overall distributions of these recorded values. Surface errors were recorded between the whole SGs (WSG) and also between the active regions (AR). Dice indices were not computed for the AR because it is not a closed surface and does not represent a volumetric region. The green bars, red bars, blue rectangles, and black I-bars denote the median, mean, one standard deviation from the mean, and the overall range of the data set, respectively. As can be seen in the figure, the wASM achieves mean dice indices of approximately 0.77. For typical structures, a Dice index of 0.8 is considered good [8]. Here, we consistently achieve Dice indices close to 0.8 for segmentation of a structure that is atypically small and lacks any contrast in the image. Mean surface errors are approximately 0.15 mm for both the WSG and the AR, which is about a half a voxel's distance in the segmented CT. Maximum surface errors are above 1 mm for the WSG but are all sub-millimetric for the AR.

Segmentations for all 5 experiments are shown color encoded with surface error in Figure 3. It can be seen that the wASM results in mean surface errors under 0.15 mm for the majority of the SG with average maximum errors of about 0.7 mm (<3 voxels). As can be seen in the figure, errors in the AR above 0.5 mm are rare and highly localized. Shown in Figure 4 are contours of a representative automatic segmentation overlaid with the CT (the volume on which segmentation was performed) and the corresponding registered μCT. It can be seen from the figure that the contours achieved by automatic segmentation of the CT are in excellent agreement with contours manually
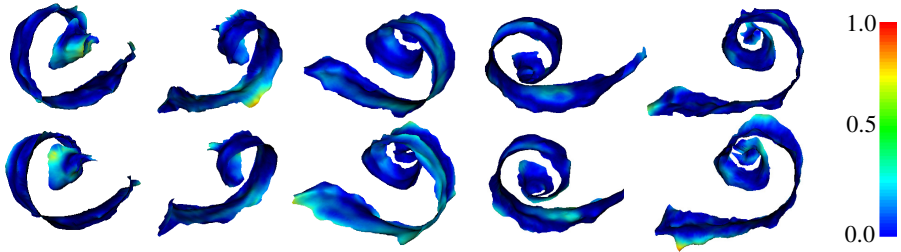
Fig. 3. Automatic (top row) and manual (bottom row) segmentations of the active region of the SG in the 5 test volumes (left-to-right) color encoded with error distance (mm)
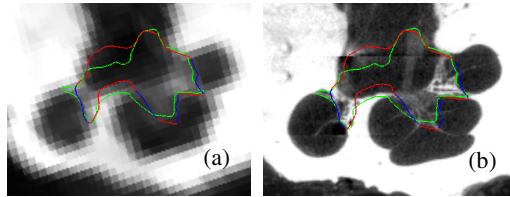
Fig. 4. Delineations of the automatic (red/blue) and manual (green) segmentation of the SG in the CT (a) and µCT (b) slice from Figure 1. The active region is shown in blue

delineated in the high resolution µCT, especially in the AR. Localization errors that are apparent in the µCT are less than 2 voxels width in the CT.

4   Conclusions

In this work, we have presented a weighted active shape based approach for identifying the SG, which lacks any contrast in conventional CT. In this approach, we rely on high resolution images of cadaveric specimens to serve two functions. First, they provide information necessary to construct an SSM of the structure, permitting segmentation of the structure in conventional imaging. Second, the high resolution images are used to validate the results. This is performed by transferring the automatically segmented structures from the conventional images to the corresponding high resolution images and comparing those structures to manual segmentations.

This work has shown that it is possible to accurately identify the location of the SG in conventional CT because the position of the SG varies predictably with respect to the cochlea. The approach we present accurately locates the SG by attracting the exterior walls of the models of intra-cochlear anatomy towards the edges of the cochlea. This approach achieves dice indices of approximately 0.77 and sub-millimetric maximum error distance in the active region, which represents the region of interest for CI stimulation. These results suggest that our segmentation approach is accurate enough to be used for position-dependent, image-guided CI programming methods.

Testing of computer assisted CI programming approaches using the presented methods has begun and has been completed on one patient thus far. The fully-automatically identified positions of the electrode and tonotopically mapped SG for

this patient are shown in Figure 1f. For that patient, the settings that were originally considered optimal by the audiologist were modified using image-based information. Improvements in the patient's hearing were statistically significant using a binomial distribution statistic for the individual speech perception metrics tested [9-10]. The patient's monosyllabic word recognition scores (a quantitative measure of hearing performance) jumped from 33% to 84%, and sentence recognition performance in noise at +10 dB signal-to-noise ratio increased from 46% to 83%. Further, the patient reported significant improvement in hearing and overall sound quality. While very preliminary, these results indicate that image-based programming, enabled by the approach described herein, may significantly improve hearing restoration for CI users.

References

1. Stakhovskaya, O., Spridhar, D., Bonham, B.H., Leake, P.A.: Frequency Map for the Human Cochlear Spiral Ganglion: Implications for Cochlear Implants. J. Assoc. Res. Otol. 8, 220–233 (2007)
2. Noble, J.H., Schuman, A., Wright, C.G., Labadie, R.F., Dawant, B.M.: Automatic Identification of Cochlear Implant Electrode Arrays for Post-Operative Assess-ment. In: Dawant, B.M., Haynor, D.R. (eds.) Medical Imaging 2011: Image Processing, Proc. of the SPIE Conf. on Med. Imag., vol. 7962, p. 796217 (2011)
3. Noble, J.H., Labadie, R.F., Majdani, O., Dawant, B.M.: Automatic segmentation of intra-cochlear anatomy in conventional CT. IEEE Trans. on Biomedical. Eng. 58(9), 2625–2632 (2011)
4. Rogers, M., Graham, J.: Robust Active Shape Model Search. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 517–530. Springer, Heidelberg (2002)
5. Green, B.F.: The orthogonal approximation of an oblique structure in factor analysis. Psychometrika 17, 429–440 (1952)
6. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active Shape Models—Their Training and Application. Comp. Vis. Image Unders. 61(1), 39–59 (1995)
7. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26, 297–302 (1945)
8. Zijdenbos, A.P., Dawant, B.M., Margolin, R.: Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. IEEE Trans. on Med. Imag. 13(4), 716–724 (1994)
9. Thornton, A.R., Raffin, M.J.: Speech-discrimination scores modeled as a binomial variable. J. Speech Hear. Res. 21, 507–518 (1978)
10. Spahr, A.J., Dorman, M.F., Litvak, L.L., Van Wie, S., Gifford, R.H., Loizou, P.C., Loiselle, L.M., Oakes, T., Cook, S.: Development and Validation of the AzBio Sentence Lists. Ear Hear. 33, 112–117 (2012)

Guiding Automatic Segmentation with Multiple Manual Segmentations

Hongzhi Wang and Paul A. Yushkevich⋆

Department of Radiology, University of Pennsylvania

Abstract. Most image segmentation algorithms are designed to esti-
mate a single segmentation for each image, where the gold standard
segmentation is often labeled by a human expert. However, it is common
that multiple manual segmentations are available for some images, e.g.
independently labeled by different experts. For efficient usages of manual
segmentations, we propose to simultaneously produce automatic estima-
tions for each expert. The key advantage of this proposal is that it al-
lows to incorporate the correlations between different experts to improve
the accuracy of automatic segmentation. In a brain image segmentation
problem, where for each image six manual segmentations are available,
we show that jointly estimating several manual segmentations produces
significant improvement over independently estimating each of them.

1 Introduction

Image segmentation is the primary mechanism for quantifying the properties of
anatomical structures and pathological formations using imaging data. Given the
often prohibitive cost of manual segmentation, accurate automatic segmentation
is highly desirable. To mimic manual segmentation, automatic segmentation is
often guided and evaluated against manual segmentations. However, segmenta-
tions labeled by different experts are often inconsistent.

Existing inconsistent manual segmentations not only reveals the significant
difficulty in performing manual segmentation, but also poses challenges on how to
develop automatic segmentation algorithms. Most automatic algorithms produce
a single solution for each image. When evaluated against inconsistent manual
segmentations, the automatic solution is either separately compared with each
of the manual segmentations or directly compared with the consensus manual
segmentation, e.g. derived by STAPLE [11]. Either way the automatic algorithm
is biased to produce solutions close to the consensus of all manual segmentations.

Employing consensus manual segmentation simplifies the evaluation process,
therefore makes the task of developing automatic methods more straightfor-
ward. However, it also sacrifices the rich information contained in the original
set of manual segmentations. Our contribution is to propose a novel scheme to

incorporate multiple manual segmentations to guide automatic segmentation. To maximize the usage of the valuable manual segmentations, we propose to simultaneously produce automatic estimations for all manual segmentations. The key advantage of this proposal is that the label correlations between different human experts can be incorporated to improve automatic segmentation.

We apply our method to segment the hippocampus in magnetic resonance images (MRI) and show significant improvement over independently producing estimations for each manual segmentation segmentation.

2 Jointly Estimating Multiple Manual Segmentations

Image segmentation can be addressed via estimating the conditional probability $p(S_F|F)$, where F is an image and S_F is a segmentation for F. Assuming that labeling different voxels is conditionally independent given the image patches located on the voxels, we have $p(S_F|F) = \prod_i p(S_F(i)|F(\mathcal{N}(i)))$, where i indexes through image voxels. $\mathcal{N}(i)$ represents a neighborhood centered at i. $F(\mathcal{N}(i))$ is the intensity patch located on the region. To estimate this probability, discriminative learning techniques learn the label distribution $p(l|F(\mathcal{N}(i)))$ from training data, e.g. [7], [8], which can be addressed by most classification algorithms. l indexes through all possible labels. The segmentation is then obtained via maximum a posterior inference, i.e. $S_F(i) = \mathrm{argmax}_l p(l|F(\mathcal{N}(i)))$.

Motivation for jointly estimating multiple manual segmentations. In the context of clinical imaging studies involving segmentation, it is common to generate repeat manual segmentations by multiple raters in order to establish inter-rater and intra-rater reliability for a manual segmentation protocol. To handle the inconsistency between multiple manual segmentations, one approach attempts to infer the "ground truth" segmentation with the consideration of the reliability of each rater [11]. However, the inferred "ground truth" loses the rich information in the original manual segmentations and the errors in deriving the hard decision of "ground truth" will affect the performance of automatic segmentation.

We advocate an alternative solution that produces a separate estimation for each manual segmentation[1]. The key advantage of this strategy is that it allows to incorporate the correlations between manual segmentations to improve the accuracy of automatic segmentation. In our experiment, we observed that some raters consistently produced larger volumes than others when segmenting the hippocampus in MRI (see section 3.1). With such correlations, observing the segmentation labeled by one rater provides meaningful information to estimate the segmentation labeled by the other. Even if only one most reliable manual segmentation is selected for one study, which is common is practice, as we show below, incorporating manual segmentations labeled by less reliable raters helps improving the automatic segmentation accuracy for the selected rater.

---

[1] A unique manual segmentation is defined as consistently labeled by one human expert in one segmentation trial.

Formulation of jointly estimating multiple manual segmentations. Jointly estimating multiple manual segmentations can be solved via estimating the following joint conditional probability $p(S_F^1, ..., S_F^m|F) \propto p(F|S_F^1, ..., S_F^m)p(S_F^1, ..., S_F^m)$

$$\propto \left[\prod_{j=1}^{m} p(F|S_F^j)\right] p(S_F^1, ..., S_F^m) \propto \left[\prod_{j=1}^{m} p(S_F^j|F)\right] p(S_F^1, ..., S_F^m) \quad (1)$$

where $S_F^1, ..., S_F^m$ estimate m manual segmentations, respectively. Given any manual segmentation for an image, we assume conditional independence between the image and any other manual segmentations for the image. The last equation is obtained by dropping the term $p(F)/p(S_F^j)$, where $p(S_F^j)$ is the prior for observing a segmentation labeled by rater j. Since it is hard to approximate this prior, we treat it as a constant and focus on optimizing the remaining terms. The first term in (1) can be estimated by separately applying discriminative learning to estimate each manual segmentation. The second term is the joint probability of observing all manual segmentations for one image, which captures their correlations. Estimating this term is difficult as well, but a good approximation can be obtained by applying pseudolikelihood [2]. We have:

$$p(S_F^1, ..., S_F^m|F) \propto \prod_{j=1}^{m} p(S_F^j|F)p(S_F^j|\{S_F^1, ..., S_F^m\}\backslash S_F^j) \quad (2)$$

In summary, each manual segmentation is estimated based on two constraints: 1) image information, which directly captures the correlation between the manual segmentation and an image; and 2) the segmentations estimated for the remaining manual segmentations, which enforces the estimated segmentations to respect the mutual correlations between different raters. As in (2), assuming assigning labels to different voxels are conditionally independent given the patches located on the voxels, we have the final approximation as $p(S_F^1, ..., S_F^m|F)$

$$\propto \prod_{j=1}^{m} \prod_{i} p(S_F^j(i)|F(\mathcal{N}(i)))p(S_F^j(i)|\{S_F^1(\mathcal{N}(i)), ..., S_F^m(\mathcal{N}(i))\}\backslash S_F^j(\mathcal{N}(i))) \quad (3)$$

2.1   Discriminative Learning

Here, we describe in detail how we estimate the conditional probabilities in (3).

Learning to approximate one manual segmentation. For each manual segmentation, to estimate $p(l|F(\mathcal{N}(i)))$, i.e. the first term in (3), we train one *segmentation classifier* using the modified AdaBoost algorithm [4],[9] for each label l to identify voxels assigned to label l in the target manual segmentation.

For better performance, we apply the *corrective learning* technique [8]. This method applies learning as an error correction tool to improving the segmentation produced by a host segmentation method. It was shown that it significantly improved the performance of the learning algorithm and the host segmentation

method [8]. In our experiments, we apply multi-atlas label fusion as the host method (see detail in section 3). Note that multi-atlas segmentation can be applied alone to estimate each manual segmentation. Applying corrective learning improves the performance. The region of interest and the features used in [8], including spatial, appearance and contextual, joint spatial-appearance and joint spatial-contextual features, are applied to train the classifiers, where the contextual features are extracted from the initial segmentation produced by the host method. To transfer the output of an AdaBoost classifier to a probability, we apply the logistic transform, i.e. $p(x) = \frac{e^x}{e^x + e^{-x}}$.

Learning the correlations between manual segmentations. To estimate the second term in (3), we train *correlation classifiers* for each manual segmentation to capture the correlation between this manual segmentation and the remaining manual segmentations. For this task, we apply spatial, contextual and joint spatial-contextual features, as in [8], to train one classifier for each label l to identify the voxels assigned to label l in the target manual segmentation. The contextual features are extracted from all the remaining manual segmentations. To effectively handle the contextual features provided by multiple manual segmentations, we merge the contextual features from different manual segmentations into one label distribution D_l^j for each label l, $D_l^j(i) = \frac{1}{m-1} \sum_{k \neq j} I(S_F^k(i) = l)$, where $I(\cdot)$ is an indicator function. The contextual features used in the correlation classifier for rater j and label l are constructed based on D_l^j.

Segmentation Algorithm. The algorithm is summarized below:

1. Given a test image F, apply a host method to produce an initial segmentation S for it. When applicable, produce one initial segmentation for each rater.
2. For $j = 1, ..., m$
   - Apply the segmentation classifier(s) learned for j^{th} manual segmentation to produce an improved estimation, S_F^j, based on image F and the initial segmentation S produced for the rater.
3. For $j = 1, ..., m$
   - Apply the correlation classifier(s) learned for j^{th} manual segmentation to update S_F^j such that (3) is maximized, i.e. selecting the label with the largest probability produced by the classifiers at each voxel.
4. If none of the automatic estimations is changed or the maximal iteration is reached, then output the estimations. Otherwise, goto 3.

3   Experiments

3.1   Imaging Data and Experiment Setup

Data and manual segmentations. For 10 images (5 controls and 5 patients with mild AD) from the Open Access Series of Imaging Studies (OASIS) [5], we produced six manual hippocampal segmentations for each image. These manual segmentations were labeled by three trained experts in two trials.

Table 1. Left: Inter-rater and intra-rater segmentation overlaps (in Dice $\frac{2|A \cap B|}{|A|+|B|}$) between the three raters. The intra-rater overlaps are between the two segmentation trails labeled by the same rater. The inter-rater overlaps are averaged over the two segmentation trials. Right: Hippocampal volume (in voxel) produced by each rater.

raters	R^1	R^2	R^3
R^1	0.902±0.020	0.872±0.024	0.847±0.032
R^2	0.872±0.024	0.915±0.020	0.846±0.043
R^3	0.847±0.032	0.846±0.043	0.836±0.046

raters	R^1	R^2	R^3
trial 1	1615±267	1787±316	1731±319
trial 2	1683±285	1811±261	1461±232

Table 1(left) summarizes the inter-rater and intra-rater reproducibility of the manual segmentation. Table 1(right) shows the hippocampal volume labeled by each rater. Note that R^2 consistently labeled larger hippocampi than R^1 in both trials. The segmentations labeled by R^3 in the second trial are significantly smaller than those produced by the same rater in the first trial. Such strong correlation can be easily seen in most individual subjects, as shown in Fig. 1.
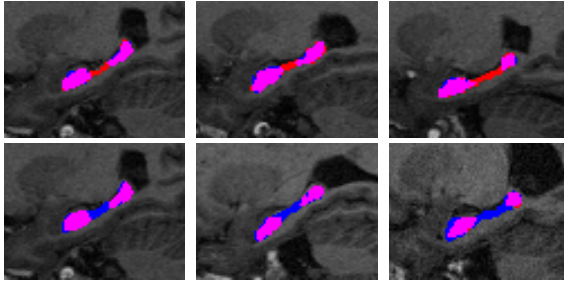
Fig. 1. Illustration of correlation between raters. First row: segmentations produced by R^1(blue) and R^2(red) in the first trial. Pink is the overlapped region. Second row: segmentations produced by R^3 in the first (blue) and second trial (red).

For each image, we derived one consensus segmentation using STAPLE [11] for the three manual segmentations produced in each trial. To incorporate the correlations between the two trials, we jointly estimate the 6 manual segmentations and the two inferred segmentations by STAPLE.

Experiment setup. For cross-validation, we randomly selected five images for training and the remaining 5 images for testing. The experiment was repeated 10 times. In each cross-validation, a different set of training and testing images were selected. The results reported below are averaged over the 10 experiments.

Details on learning segmentation classifiers. Since the state-of-the-art hippocampus segmentation are all produced by multi-atlas label fusion (MALF), e.g. [6],[3],[10], we applied MALF as the host segmentation method to produce the
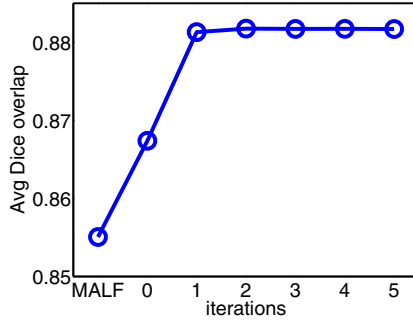
Fig. 2. Segmentation accuracy (in Dice) at each iteration for all raters. The results are averaged over 10 cross-validation experiments. The performance of independently applying error correction to estimate each manual segmentation is given at iteration 0.

initial hippocampus segmentation for corrective learning to learn the segmentation classifier for each manual segmentation.

Through deformable registration, MALF warps multiple atlases, i.e. pre-labeled images, to a target image, and uses a "label fusion" strategy to derive a consensus segmentation. To implement MALF, image guided registration is performed by the Symmetric Normalization (SyN) algorithm implemented by ANTS [1] between each pair of the atlas image, i.e. the training image, and the test image. For label fusion, we apply image similarity based local weighted voting technique, which is shown to be the most effective label fusion techniques in recent studies [6], [10]. The voting weights were computed based on image patches of size $5 \times 5 \times 5$ by using the joint label fusion algorithm [10].

To produce the initial segmentation used in corrective learning for one manual segmentation, we use the segmentation labeled by the corresponding rater to define the atlas. For each cross-validation, we also apply MALF to produce an initial segmentation for each training image by using the remaining training images as atlases and use the segmentation produced by MALF for training images to train the segmentation classifiers for each manual segmentation. For each cross validation, learning all segmentation classifiers and all correlation classifiers took about 1 hour and 30 minutes on a 2GHz CPU, respectively.

3.2   Results

Convergence. Fig. 2 shows the average segmentation performance produced by MALF, MALF + corrective learning (iteration 0), and our joint segmentation algorithm at different iterations. Typically, the iterative optimization converges within only a few iterations, with the first iteration producing the maximal performance improvement and dramatic diminishing performance gains in later iterations. In our experiment, we set the maximal iteration to be 10.

Quantitative comparison. Table 2 compares the performance between our method with separately estimating each manual segmentation. Corrective learning

substantially improved the accuracy produced by MALF. Our method further improved the accuracy to the level greater than inter-rater accuracy for each rater. The improvements for each rater are statistically significant, with $p < 0.001$ on the paired Students t-test. Fig. 3 shows some segmentation results produced by applying corrective learning alone and by our method, respectively.

Table 2. Segmentation accuracy (in Dice) with respect to each manual segmentation. R_k^j is the segmentation produced by rater R^j in the k_{th} segmentation trial. STP_k is the consensus segmentation produced by STAPLE for the k_{th} segmentation trial.

rater	MALF	MALF+learning	MALF+jointSeg
R_1^1	0.865±0.021	0.878±0.014	**0.890**±0.015
R_1^2	0.852±0.020	0.869±0.018	**0.881**±0.018
R_1^3	0.833±0.032	0.837±0.025	**0.859**±0.022
STP_1	0.869±0.023	0.888±0.016	**0.900**±0.014
R_2^1	0.859±0.023	0.864±0.018	**0.880**±0.021
R_2^2	0.861±0.018	0.877±0.017	**0.887**±0.016
R_2^3	0.829±0.035	0.840±0.024	**0.857**±0.023
STP_2	0.871±0.022	0.886±0.016	**0.900**±0.017

image   MALF   MALF+learning MALF+JointSeg
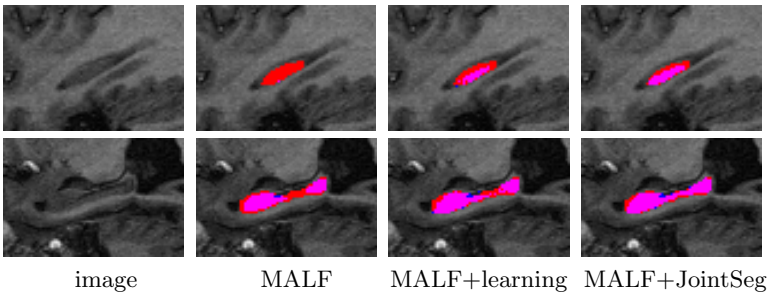
Fig. 3. Sagittal views of hippocampus segmentation results. Red: one of the manual segmentations for the image; Blue: automatic segmentation; Pink: overlap between manual and automatic.

Our results compare well to the state-of-the-art hippocampus segmentation performance. For example, [6] reported average ∼0.87 (Dice) for hippocampus using 29 atlases. [3] reported average 0.887 (Dice) using 79 atlases. Our final results for R^1 and R^2 are >0.880 (Dice)[2], but we only used 5 training images, which is only a small fraction of those used by the competing work.

4   Conclusion

As an important evaluation target, manual segmentation is crucial in the development of automatic segmentation algorithms. We developed a technique to

---

[2] Our results for R^3 are lower due to the poor intra-rater performance.

incorporate multiple inconsistent manual segmentations to improve the performance of automatic segmentation. Via experiments on hippocampus segmentation in MRI, we showed the advantage of our method over traditional approaches. Note that including the segmentations produced by less reliable raters helped to better estimate the segmentations by more reliable raters. Our work offers a new perspective on how to more effectively use the valuable manual segmentations.

References

1. Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12(1), 26–41 (2008)
2. Besag, J.: Statistical analysis of non-lattice data. J. R. Statist. Soc. B 24(3), 179–195 (1975)
3. Collins, D., Pruessner, J.: Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. NeuroImage 52(4), 1355–1366 (2010)
4. Freund, Y., Schapire, R.: A Decision-theoretic Generalization of On-line Learning and an Application to Boosting. In: Vitányi, P.M.B. (ed.) EuroCOLT 1995. LNCS, vol. 904, pp. 23–27. Springer, Heidelberg (1995)
5. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults. Journal of Cognitive Neuroscience 22(12), 2677–2684 (2010)
6. Sabuncu, M., Yeo, B., Leemput, K.V., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. IEEE TMI 29(10), 1714–1720 (2010)
7. Tu, Z., Zheng, S., Yuille, A., Reiss, A., Dutton, R., Lee, A., Galaburda, A., Dinov, I., Thompson, P., Toga, A.: Automated extraction of the cortical sulci based on a supervised learning approach. IEEE TMI 26(4), 541–552 (2007)
8. Wang, H., Das, S., Suh, J.W., Altinay, M., Pluta, J., Craige, C., Avants, B., Yushkevich, P.: A learning-based wrapper method to correct systematic errors in automatic image segmentation: Consistently improved performance in hippocampus, cortex and brain segmentation. NeuroImage 55(3), 968–985 (2011)
9. Wang, H., Suh, J.W., Das, S., Pluta, J., Altinay, M., Yushkevich, P.: Hippocampus segmentation using a stable maximum likelihood classifier ensemble algorithm. In: 2011 IEEE International Symposium on Proceeding of: Biomedical Imaging: From Nano to Macro (2011)
10. Wang, H., Suh, J., Pluta, J., Altinay, M., Yushkevich, P.: Optimal Weights for Multi-atlas Label Fusion. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 73–84. Springer, Heidelberg (2011)
11. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE TMI 23(7), 903–921 (2004)

Atlas-Based Probabilistic Fibroglandular Tissue Segmentation in Breast MRI

Shandong Wu[1], Susan Weinstein[2], and Despina Kontos[1]

[1] Computational Breast Imaging Group, Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[2] Breast Imaging Section, Department of Radiology,
Hospital of the University of Pennsylvania, Philadelphia, PA, USA
`shandong.wu@uphs.upenn.edu`

Abstract. In this paper we propose an atlas-aided probabilistic model-based segmentation method for estimating the fibroglandular tissue in breast MRI, where a novel fibroglandular tissue atlas is learned to aid the segmentation. The atlas represents a pixel-wise likelihood of being fibroglandular tissue in the breast, which is derived by combining deformable image warping, using aligned breast contour points as landmarks, with a kernel density estimation technique. A mixture multivariate model is learned to characterize the breast tissue using MR image features, and the segmentation is subsequently based on examining the posterior probability where the learned atlas is incorporated as the prior probability. In our experiments, the algorithm-generated segmentation results of 10 cases are compared to the manual segmentations, verified by an experienced breast imaging radiologist, to assess the accuracy of the algorithm, where the Dice's Similarity Coefficient (DSC) shows a 0.85 agreement. The proposed automated segmentation method could be used to estimate the volumetric amount of fibroglandular tissue in the breast for breast cancer risk estimation.

1 Introduction

Breast magnetic resonance imaging (MRI) provides 3D scanning and has emerged as an effective modality for breast cancer risk assessment for high-risk population [1, 2]. Studies indicate that the percentage of fibroglandular tissue (FT%) computed in breast MRI is related to breast cancer risk [2, 3]. To estimate the FT% in breast MRI, accurate segmentation of the fibroglandular tissue from the breast is a fundamental step. Fibroglandular tissue segmentation is challenging in several aspects. First, segmenting the breast as an organ from the remaining parts of the MR image is critical [4]. Second, fibroglandular tissue may be present anywhere in the breast with varying amounts, shapes, and patterns, which is challenging to model by simple geometric descriptors. In addition, the bias field is common in breast MRI where the intensity inhomogeneity may considerably affect the appearance of tissue properties.

The problem of automated fibroglandular tissue segmentation has received little attention in the literature to date [4, 5]. In addition to the qualitative estimation of the fibroglandular tissue by visual assessment [3], most previous studies rely on semi-automated segmentation methods such as interactive thresholding [1] or clustering [2, 6].

Fuzzy C-means (FCM) has been used where usually the number of clusters either relies on user's determination [2] or based on intensity range assumptions followed by interactive adjustments [6]. Visual assessment and interactive methods are subjective and introduce inter- and intra-reader variability [1-3, 6]. An automatic method is reported in [4] to simultaneously segment heart, pectoral, lung, and breast, in which the fibroglandular tissue is also processed in the axial view of breast MRI. In dealing with intensity inhomogeneity, FCM, N3 [7], and CLIC [8] have been recently tested for breast MRI [5], where segmentation performance of different combinations of these algorithms is reported based mainly on the visual evaluation by radiologists.

We propose an atlas-aided probabilistic model-based method for fibroglandular tissue segmentation, where a likelihood atlas is constructed to aid the segmentation. The breast tissue is characterized by a mixture model and segmentation is based on posterior probability evaluation by incorporating the likelihood atlas as a prior. Our major contributions include i) constructing the novel fibroglandular tissue likelihood atlas and ii) proposing an atlas-aided probabilistic model-based method to address the challenging problem of fibroglandular tissue segmentation in breast MRI.

2   Methods

The proposed segmentation method includes two preprocessing steps. First, the outline of the breast is segmented in the breast MR images, implemented based on [9]. This step is critical to the fibroglandular tissue estimation because it precludes the interferences coming from the non-breast regions. Second, we further process the segmented breast by applying the nonparametric non-uniform intensity normalization (N3) algorithm [7] to reduce bias field. After that, our atlas-aided model-based segmentation method consists of three major modules, i.e., 1) likelihood atlas construction, 2) model learning, and 3) segmentation, as described in the following.

2.1   Fibroglandular Likelihood Atlas Learning

We construct a fibroglandular tissue likelihood atlas to aid the segmentation. The atlas represents a statistical likelihood of being fibroglandular tissue for each pixel on a standard breast template and it is derived by deformable breast image warping and kernel density estimation techniques. Our atlas is similar in principle to shape/organ atlases in the notion of "statistical atlas" but differs in that it describes the likelihood of breast tissue type rather than the anatomy/structure of a shape or organs [4, 10]. In our approach, we choose not to treat fibroglandular tissue as a well-defined structure in the image (e.g., the shape of the fibroglandular tissue can vary largely between cases), but rather model its spatial distribution to construct the likelihood atlas.

The likelihood atlas is derived based on "ground truth" segmented fibroglandular tissue (e.g., provided by an expert), from a set of training samples of 3D breast MRI scans, and processing is performed separately at the intra-case and inter-case levels.

Intra-case Learning: For a given 3D scan, the first step is to create a case-specific 2D breast template in the form of a closed contour of the breast. To do so, the slices of the given scan are aligned via rigid registration [11] to the center slice. Then, the breast contour for each slice is extracted [9] and interpolated into equidistantly distributed points of a fixed number (200 in our settings) (Fig. 1(a)(b)). The points of

breast contours across the different slices are then matched by the non-linear dynamic time warping (DTW) algorithm [12] to establish point-correspondence (Fig. 1(c)), from which a breast contour is produced by averaging the corresponding matched-points across different slices. The averaged contour delineates a 2D mean shape of the breast defined as the case-specific 2D breast template for the given scan (Fig. 1(c)).
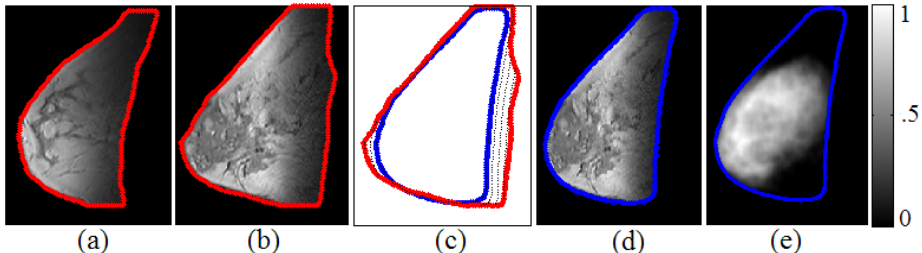
Fig. 1. Intra-case fibroglandular likelihood atlas construction. (a) (b) Two representative slices of a 3D MRI scan where the breast contours are plotted in red. (c) The DTW-based matching of the breast contours between (b) (red) and the generated case-specific 2D breast template (blue). (d) Based on the built point-correspondence in (c), the breast in (b) is deformed to the breast template (blue). (e) The learned case-specific 2D likelihood atlas where the colorbar indicates pixel-wise likelihood of being fibroglandular tissue within the breast template.

Next, pixel-wise likelihood is learned for each point in the case-specific 2D breast template. For each of the training slice, we take its breast contour points as original landmarks and warp the whole slice to align it with the case-specific breast template whose breast contour points are taken as the corresponding target landmarks (Fig. 1(d)). Here the warping is implemented in a deformable manner by the Delaunay triangulation and cubic interpolation [13]. In this way we transfer all slices and the associated segmented fibroglandular tissues into the same coordinate space. Subsequently, we project the segmented fibroglandular tissues for each slice into the case-specific breast template space where a 2D spatial distribution of the accumulated projected points is formed. Let $\{Z_i\}_{i=1}^{N}$ represent the N accumulated points where $Z_i = [x_i, y_i]$ encodes the 2D coordinates for point i, we apply the following 2D kernel density estimation [14] to derive a continuous distribution function,

$$f_H(Z) = \frac{1}{N}\sum_{i=1}^{N} G_H(Z - Z_i) \ , \tag{1}$$

where Gaussian kernel is selected for G_H and the bandwidth H is automatically determined according to [14]. Function $f_H(Z)$ represens a pixel-wise likelihood of being fibroglandular tissue in the breast template, which yields the case-specific fibroglandular likelihood atlas for the given MRI scan (Fig. 1(e)).

Inter-case Learning: By repeating the above intra-case processing for multiple 3D scans we obtain multiple case-specific likelihood atlases (Fig. 2(a)-(c)). At the inter-case level, we apply similar processes as described above to generate the likelihood atlas over the entire training set. First, an overall 2D breast template is generated from averaging the DTW-matched contour points extracted from multiple case-specific 2D atlases. Then all the case-specific atlases are deformed by the Delaunay triangulation

and cubic interpolation [13] to the overall breast template space to average out an overall fibroglandular likelihood atlas (denoted by A_1; Fig. 2(d)). Since a breast primarily consists of fibroglandular tissue and fat, a fat likelihood atlas A_2 can be obtained simultaneously as $A_2 = 1 - A_1$. The two atlases will be used in Section 2.3.
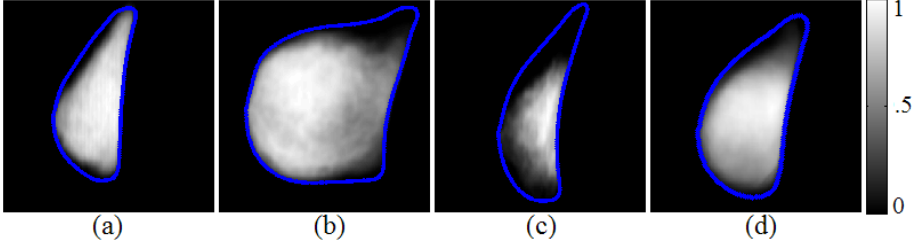
Fig. 2. Inter-case fibroglandular likelihood atlas construction. Three case-specific likelihood atlases derived from three different cases are shown in (a), (b), and (c), respectively, and (d) shows an overall atlas (denoted by A_1) learned over 9 different cases. The colorbar for (a)-(d) indicates pixel-wise likelihood of being fibroglandular tissue.

2.2　Fibroglandular Tissue Model Learning

A mixture model of Gaussian distribution is employed to model the fibroglandular tissue and fat, respectively, using image-derived features of the tissue appearance. For more efficient model learning, we choose to work at the superpixel [15] rather than pixel level (note here a pixel essentially represents a voxel). We adapt a publicly available code in [15] to over-segment to a set of superpixels (superpixel number is initially set to 300 by referring to [16]) in which the superpixels in the non-breast regions are filtered out (Fig. 3(a)). Then tissue feature vectors (X) are computed at the superpixel level including, mean intensity, intensity variance, skewness, and kurtosis [17]. For each class of tissue, the following multivariate mixture model Θ is learned,

$$P(X \mid \Theta) = \sum_{k=1}^{K} \omega_k \cdot \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \cdot e^{[-1/2 \cdot (X-\mu_k)^T \Sigma_k^{-1}(X-\mu_k)]} , \qquad (2)$$

where $d = 4$, ω_k, μ_k and Σ_k represent the weight, mean, and covariance matrix of the k-th Gaussian component, respectively. The integrated expectation–maximization (EM) and iterative pairwise replacement algorithm (IPRA) [12] are adapted to learn the model, in which the mixing component number K is automatically determined. Briefly, the EM algorithm performs an initial estimation of the components and a refinement process is followed by the IPRA algorithm. The learned two models are denoted by $\Theta_m = \{\{\omega_k, \mathcal{N}(\mu_k, \Sigma_k)\}_{k=1}^{K}\}_{m=1}^{2}$ where Θ_1 is for fibroglandular tissue and Θ_2 for fat, respectively.

2.3　Fibroglandular Tissue Segmentation

We formulate the segmentation of a test breast (i.e., the set of 2D slices) as the atlas-aided posterior probability assessment after incorporating the likelihood atlases as prior probability. To do so, the correspondence between each of the test slices (Fig. 3(a)) and the atlas (i.e., the breast template; Fig. 3(b)) needs to be established. We utilize a similar warping process as described in the atlas construction to deform the likelihood atlas to

align with a test slice by the Delaunay triangulation and cubic interpolation [13] (Fig. 3(c)). For segmenting a single test slice, we compute pixel-wise posterior probability. Note that the atlas gives pixel-wise likelihood but the test image features are extracted in terms of superpixels. Therefore, in computing the pixel-wise posterior probability for each pixel within a superpixel region, the same set of image features is repeatedly used for each pixel belonging to that superpixel (refer to equation (4)). This way the segmentation can benefit both from the robustness of superpixels to noise and the sensitivity to small blobs of fibroglandular tissue. For each pixel q of a superpixel S in the test slice I (assume I is a gray level image and $I(q) \in [0,255]$), the posterior probability (Fig. 3(f)) w. r. t the learned models $\{\Theta_m\}_{m=1}^2$ is computed by:

$$P(\Theta_m \mid q) \propto P(q \mid \Theta_m) \cdot P(\Theta_m(q)) \ , \tag{3}$$

where $P(q \mid \Theta_m)$ (Fig. 3(e)) is calculated at its superpixel level using equation (2):

$$P(q \mid \Theta_m) = P(S \mid \Theta_m) = [-\log \sum_{k=1}^K \omega_k \cdot \mathcal{N}(S; \mu_k, \Sigma_k)]^{-1} \ , \tag{4}$$

and $P(\Theta_m(q))$ (Fig.3d) is based on atlas A_m weighted by a regularization item R_m:

$$P(\Theta_m(q)) = R_m(q) \cdot A_m(q), \text{ where } R_1(q) = 1 - \frac{I(q)}{255}; R_2(q) = 1 - R_1(q). \tag{5}$$

In equation (5), $A_m(q)$ is the corresponding likelihood value of the aligned atlas and $R_m(q)$ (R_1 for fibroglandular tissue and R_2 for fat) is a regularization/weighting
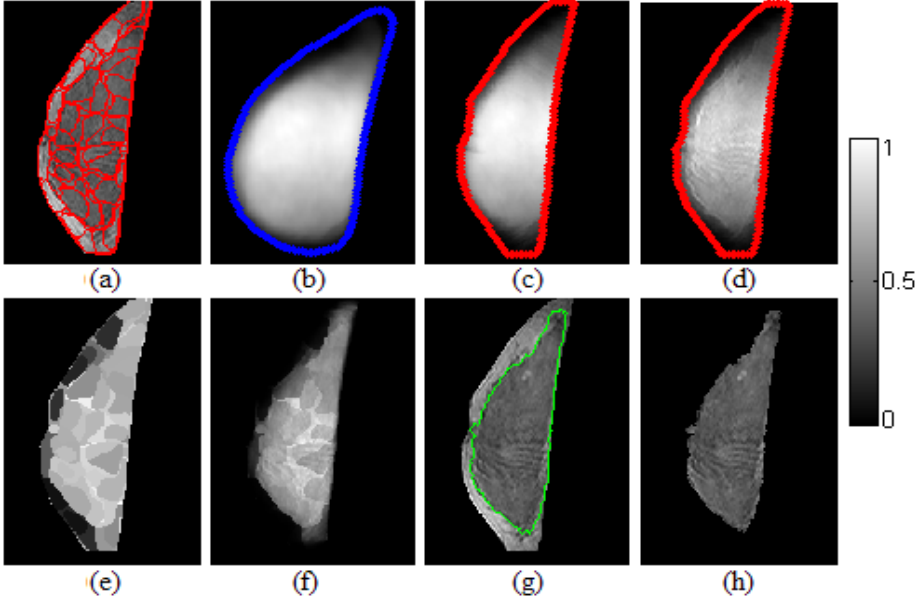
Fig. 3. Fibroglandular tissue segmentation by examining posterior probability. The colorbar is for (b)-(f) and indicates the likelihood of being fibroglandular tissue. (a) Test slice in terms of superpixel representation. (b) A training-set learned fibroglandular likelihood atlas (A_1). (c) The atlas in (b) is warped to align with the test slice. (d) Regularized atlas ($P(\Theta_1(q))$). (e) Superpixel-wise probability profile of $P(S \mid \Theta_1)$. (f) Pixel-wise posterior probability profile $P(\Theta_1 \mid q)$. (g) Segmentation contour. (h) Segmented fibroglandular tissue.

coefficient based on the intensity properties of the test slice. The term R_m is considered as a further regularization to the magnitude of the a priori learned atlas to better fit the given test case for improved segmentation. Fibroglandular tissue is segmented (Fig. 3(g)(h)) as the pixels that satisfy $P(\Theta_1 \mid q) > P(\Theta_2 \mid q)$.

3   Results

In our experiments we use 10 3D bilateral MRI cases randomly selected from a high-risk screening population [18], with cancer-unaffected, T1-weighted, non-fat-suppressed imaging in the sagittal view. Three cases fall in the Breast Imaging-Reporting and Data System (BI-RADS) density category III and 7 in category IV. There are 56 slices for each scan, resulting in a total of 10×56=560 2D MRI slices used in our experiments. Women were imaged prone in a 1.5T scanner with dedicated surface breast coil; matrix size: 256×256; slice thickness: 2-3.5mm; flip angle: 20°. The fibroglandular tissue for each case (after bias correction) is manually segmented by a 10-year experienced breast imaging radiologist, who estimates the fibroglandular regions in the breast by adapting a previously-suggested intensity thresholding approach [1]. The radiologist has the option to first select a rough region of fibroglandular tissue by outlining a closed contour prior to operating the thresholding. This manual segmentation is considered as our ground truth for validation.

Our experiments are based on leave-one-out cross validation: in each loop 9 cases are used for model training and atlas learning and the left one is used for testing. The segmentation accuracy is based on assessing agreement of the 3D volume between the algorithm- and manual-generated segmentation by the Dice's Similarity Coefficient (DSC). We also compare our segmentation against the intensity-based FCM clustering algorithm [19], which has previously been used for fibroglandular tissue segmentation in breast MRI. In our implementation we also apply the two preprocessing steps (e.g., the breast boundary segmentation and bias field correction) and we fix the number of cluster to 2 to guide the FCM to divide the breast into two broad intensity-based clusters: fibroglandular tissue and fat. Finally, we also compare our proposed method to a conventional (e.g., without atlas prior) Gaussian mixture model (GMM) where $P(\Theta_m)$ is set to 0.5 for equal occurrence frequency for the two breast tissue classes.

Table 1. Segmentation performance (DSC) comparison for the 10 cases

Case	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	Average
Atlas-GMM	0.80	0.81	0.82	0.89	0.88	0.87	0.88	0.83	0.84	0.85	**0.85**
GMM	0.57	0.47	0.62	0.79	0.73	0.71	0.79	0.74	0.70	0.77	**0.69**
FCM	0.66	0.68	0.73	0.87	0.79	0.80	0.82	0.70	0.79	0.79	**0.76**

Table 1 lists the volumetric DSC performance comparison for each of the 10 cases. Overall we achieved an average segmentation accuracy of DSC=0.85. As can be seen, our proposed atlas-aided method brings a significant improvement in segmentation accuracy over FCM and the conventional GMM method (the corresponding p values

are 0.0009 and 0.004 for comparing our atlas-aided method against the GMM and FCM, respectively, using a paired t-test), which suggests a benefit from incorporating the constructed fibroglandular tissue likelihood atlases in the segmentation process. Segmentation examples are shown in Fig. 4 with the comparison to the corresponding manual segmentations.

As a preliminary validation of the dependency of our method on the pre-selected number of superpixels, we also experiment with initializing the superpixel algorithm [15] by 1000 superpixels. We observe that the atlas-aided segmentation accuracy remains relatively stable, with a slight increase of ~ 0.3% relative to 300 superpixels.
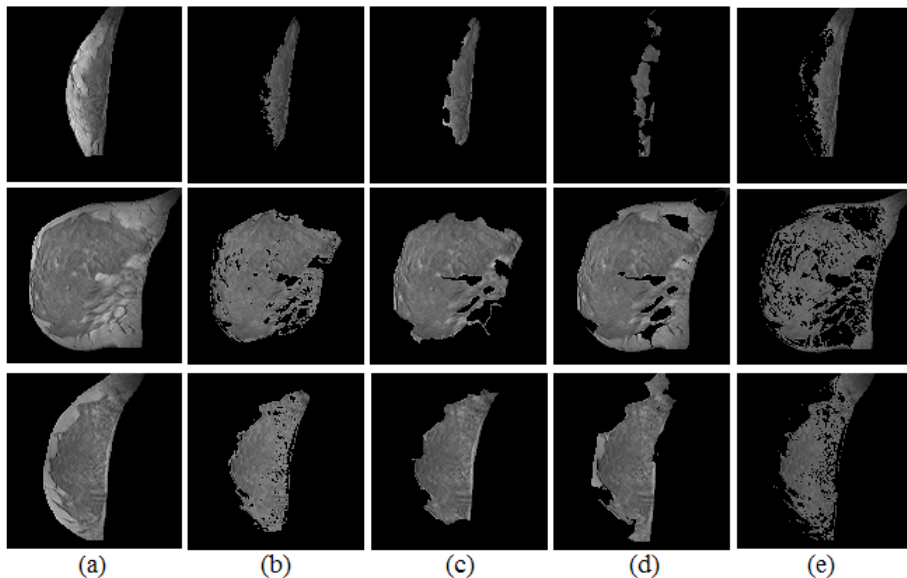
Fig. 4. Segmentation examples and the comparison to manual segmentation. Each row shows one case example. (a) Segmented and bias corrected breast. (b) Manual segmentation. (c) Atlas-aided GMM segmentation. (d) GMM segmentation (without atlas). (e) FCM segmentation.

4 Discussion and Conclusion

We propose an atlas-aided probabilistic model-based approach for fibroglandular tissue segmentation in breast MRI. The constructed statistical atlases serve as an effective prior of fibroglandular tissue likelihood in posterior probability assessment for segmentation. Experiment results demonstrate superior segmentation accuracy for our approach, compared to other commonly used approaches, suggesting a benefit of incorporating the constructed atlases in the proposed segmentation method. This work is performed to establish proof-of-concept and feasibility for our algorithm. Based on the segmentation, it is straightforward to derive a volumetric percentage of the fibroglandular tissue (FT%), which could ultimately be used for breast cancer risk estimation. In addition, the computed posterior probabilities could essentially be

translated into voxel-wise percentages of fibroglandular tissue to alleviate partial volume effects [2] and provide aggregated relative measures of fibroglandular tissue.

Future work will include further validation with a larger dataset including different vendors and protocols, extending the atlas construction in the 3D space, and employing additional image features (e.g., textures) to characterize the tissue properties. Multi-reader variability study on the manual segmentation will also be considered. Further testing of performance dependence on the number of superpixels is also warranted to fully validate our method's robustness.

References

1. Khazen, M., Warren, R.M., et al.: A pilot study of compositional analysis of the breast and estimation of breast mammographic density using three-dimensional T1-weighted magnetic resonance imaging. Cancer Epidemiol. Biomarkers Prev. 17, 2268–2274 (2008)
2. Klifa, C., Carballido-Gamio, J., Wilmes, L., Laprie, A., Shepherd, J., Gibbs, J., Fan, B., Noworolski, S., Hylton, N.: Magnetic resonance imaging for secondary assessment of breast density in a high-risk cohort. Magn. Reson. Imag. 28(1), 8–15 (2010)
3. King, V., Brooks, J.D., Bernstein, J.L., Reiner, A.S., Pike, M.C., Morris, E.A.: Background parenchymal enhancement at breast MR imaging and breast cancer risk. Radiology 260(1), 50–60 (2011)
4. Mérida, A.G., Kallenberg, M., Martí, R., Karssemeijer, N.: Fully automatic fibroglandular tissue segmentation in breast MRI: atlas-based approach. In: MICCAI Workshop (2011)
5. Lin, M., Chan, S., Chen, J.H., Chang, D., Nie, K., Chen, S.T., Lin, C.J., Shih, T.C., Nalcioglu, O., Su, M.Y.: A new bias field correction method combining N3 and FCM for improved segmentation of breast density on MRI. Medical Physics 38(1) (2011)
6. Nie, K., Chen, J.H., Chan, S., Chau, M.K., Yu, H.J., Bahri, S., Tseng, T., Nalcioglu, O., Su, M.Y.: Development of a quantitative method for analysis of breast density based on three-dimensional breast MRI. Med. Phys. 35, 5253–5262 (2008)
7. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE TMI 17, 87–97 (1998)
8. Li, C., Xu, C., Anderson, A.W., Gore, J.C.: MRI Tissue Classification and Bias Field Estimation Based on Coherent Local Intensity Clustering: A Unified Energy Minimization Framework. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 288–299. Springer, Heidelberg (2009)
9. Wu, S.D., Weinstein, S.P., Conant, E.F., Localio, A.R., Schnall, M.D., Kontos, D.: Fully automated chest wall line segmentation in breast MRI by using context information. In: SPIE Medical Imaging: CAD (2012)
10. Chintalapani, G., Ellingsen, L.M., Sadowsky, O., Prince, J.L., Taylor, R.H.: Statistical Atlases of Bone Anatomy: Construction, Iterative Improvement and Validation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 499–506. Springer, Heidelberg (2007)

11. Bergen, J.R., Anandan, P., Hanna, T.J., Hingorani, R.: Hierarchical Model-based Motion Estimation. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 237–252. Springer, Heidelberg (1992)
12. Wu, S.D., Li, Y.F.: Flexible signature descriptions for adaptive motion trajectory representation, perception and recognition. Pattern Recognition 42, 194–214 (2009)
13. Watson, D.F.: Contouring: A guide to the analysis and display of spacial data. Perg. (1994)
14. Botev, Z.I., Grotowski, J.F., Kroese, D.P.: Kernel density estimation via diffusion. Annals of Statistics 38(5), 2916–2957 (2010)
15. Mori, G.: Guiding model search using segmentation. In: IEEE ICCV (2005)
16. Yang, A.Y., Wright, J., Ma, Y., Sastry, S.S.: Unsupervised segmentation of natural images via lossy data compression, Com. Vis. and Imag. Und. 110, 212–225 (2008)
17. Byng, J.W., Boyd, N.F., Fishell, E., Jong, R.A., Yaffe, M.J.: Automated analysis of mammographic densities. Physics in Medicine and Biology 41, 909–923 (1996)
18. Weinstein, S.P., Localio, A.R., Conant, et al.: Multimodality screening of high-risk women: a prospective cohort study. J. Clin. Onco. 27(36), 6124–6128 (2009)
19. Xing, Y., Xue, Z., Englander, S., Schnall, M.D., Shen, D.: Improving Parenchyma Segmentation by Simultaneous Estimation of Tissue Property T1 Map and Group-Wise Registration of Inversion Recovery MR Breast Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 342–350. Springer, Heidelberg (2008)

Fast 3D Spine Reconstruction of Postoperative Patients Using a Multilevel Statistical Model

Fabian Lecron[1], Jonathan Boisvert[2], Saïd Mahmoudi[1],
Hubert Labelle[3], and Mohammed Benjelloun[1]

[1] Computer Science Dept., Faculty of Engineering, University of Mons, Belgium
`Fabian.Lecron@umons.ac.be`
[2] Information and Communications Technologies,
National Research Council, Canada
`Jonathan.Boisvert@nrc-cnrc.gc.ca`
[3] Sainte-Justine Hospital, Montréal, Canada

Abstract. Severe cases of spinal deformities such as scoliosis are usually treated by a surgery where instrumentation (hooks, screws and rods) is installed to the spine to correct deformities. Even if the purpose is to obtain a normal spine curve, the result is often straighter than normal. In this paper, we propose a fast statistical reconstruction algorithm based on a general model which can deal with such instrumented spines. To this end, we present the concept of multilevel statistical model where the data are decomposed into a within-group and a between-group component. The reconstruction procedure is formulated as a second-order cone program which can be solved very fast (few tenths of a second). Reconstruction errors were evaluated on real patient data and results showed that multilevel modeling allows better 3D reconstruction than classical models.

Keywords: 3D reconstruction, spine, statistical shape model, multilevel modeling.

1 Introduction

Three-dimensional reconstruction of the spine is a valuable process to study spinal deformities such as scoliosis. It allows to determine clinical indices helping diagnosis and treatment. It normally needs to be performed based on radiographs to allow a natural standing position for the patient and to reduce as much as possible the exposition of young patients to ionizing radiations.

Usual approaches to reconstruct the spine from two radiographs consists in manually identifying corresponding landmarks on the views and matching them in a three-dimensional space. These methods required to locate at least six points per vertebra [1]. Other authors proposed to increase the number of points to be located by considering landmarks that are only visible in one radiograph [8]. These methods are very time-consuming and are hardly transposable to a medical practice.

In order to reduce the amount of manual intervention, statistical reconstructions methods later appeared. In those methods, a reduced set of input is provided by the user and the rest of the model is inferred with the help of a statistical shape model. For example, Humbert *et al.* [5] proposed to evaluate a parametric model based on the spinal centerline, Moura *et al.* [9] inferred an articulated model of the spine based on splines, and Boisvert *et al.* [2] formulated the estimation of the spine shape as a second-order cone program.

Patients with spinal deformities can be treated in various ways depending on the severity of the deformation. For severe cases, surgery can be recommended. All the methods we have just described are valuable, but are always focused on patients who have not undergone any surgery. Surgical treatment of scoliosis consists in applying instrumentation to the spine in order to redress the spine and maintain the correction. Even if the purpose is to obtain a normal spine curve, the result is often neither a normal spinal curve nor a scoliotic one (see Fig. 1 for instance). Therefore, it is difficult to capture these specific deformations with a classical statistical model. In this context, we propose to use a more general model adapted to hierarchical structures like the spine: a multilevel statistical shape model. The advantage of such a model is to represent the dependency between one vertebra and the others. As a result, several sub-models are built and can be treated separately, each level characterizing one sub-model. In the literature, the inter-vertebra dependence between pairs of vertebræ has already been modeled in [3]. Our multilevel framework is however more generic since various group structures can be selected. We can, for instance, represent the dependence between individual vertebra, between duos of vertebræ, between triplets, *etc.*

To the best of our knowledge, this paper is the first report of an interactive and fast 3D reconstruction method of the spine when surgical instrumentation is present. Furthermore, our approach introduces the use of the multilevel statistical shape modeling to the problem of 3D shape reconstruction. We will show that our method provides better results than classical statistical models.

2 Method

2.1 Multilevel Statistical Shape Model

While principal component analysis (PCA) is usually required to build a statistical shape model, multilevel component analysis (MCA) is the basis to design a multilevel statistical shape model. The concept of MCA was introduced in [10] as an extension of PCA for hierarchical structures. If we consider a model with 2 levels, the idea is to decompose the data into a within-individual and a between-individual component. Let us assume a sample with N items, divided into K groups of size K_k. An item i belonging to the group k according to the variable j is denoted by: x_{ijk}, with $i \in [1, \ldots, K_k]$, $k \in [1, \ldots, K]$, and $j \in [1, \ldots, J]$. Based on the Cronbach and Webb's model [4], x_{ijk} can be decomposed into a within-group and a between-group term, such as:
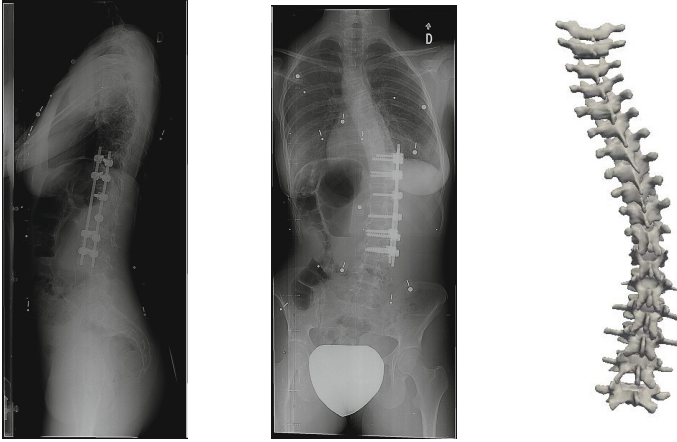
Fig. 1. Radiographs of a postoperative patient and the associated reconstruction. Left: Lateral view. Center: Postero-anterior view. Right: 3D reconstruction based on our approach.

$$x_{ijk} = x_{\bullet j \bullet} + (x_{\bullet jk} - x_{\bullet j \bullet}) + (x_{ijk} - x_{\bullet jk}), \tag{1}$$

where $x_{\bullet j \bullet} = \frac{1}{N} \sum_{k=1}^{K} \sum_{i=1}^{K_k} x_{ijk}$, and $x_{\bullet jk} = \frac{1}{K_k} \sum_{i=1}^{K_k} x_{ijk}$. In the relation (1), $(x_{\bullet jk} - x_{\bullet j \bullet})$ is the between-group term, while $(x_{ijk} - x_{\bullet jk})$ is the within-group one.

Based on the decomposition of the equation (1), a multilevel model is defined as several sub-models that can be treated separately. Let us consider the spine as a hierarchical structure such as in Fig. 2.

Let us assume a sample of I patients characterized by K vertebræ. In [7], the authors proposed a multilevel modelization of the vertebræ. Here, we develop a deformable model that can represent all the spine. As a consequence, in the
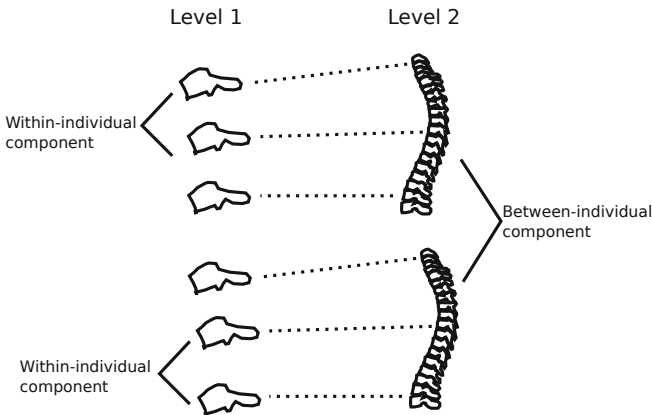
Fig. 2. Multilevel representation of the spine for a sample of patients

proposed modelization, each patient of the hierarchical structure can be viewed as a group at the first level. As a consequence, the within-group part of the model represents the variability between vertebræ while the between-group part concerns the inter-patient variability. If we assume that a patient (a spine) is represented according to J variables, a deformable model of a given vertebra x_i with $i \in [1, \ldots, K]$ is defined by:

$$x_i = \bar{x} + \phi_{W,i} d_W + \phi_B d_B, \tag{2}$$

where $\bar{x}$ is a column-vector of size J standing for the mean computed over all the objects in the sample, $\phi_{W,i}$ denotes the lines of ϕ_W, a $J \times R_W$ matrix containing the within-group principal components, d_W is a $R_W \times 1$ vector representing the weights controlling the deformation of the within-group term, ϕ_B stands for the $J \times R_B$ matrix representing the between-group principal components, d_B is the $R_B \times 1$ vector of weights controlling the deformation of the between-group term.

The interest of MCA is that the parameters of the equation (2) can be determined separately (the demonstration can be found in [10]).

Let X, be a $N \times J$ matrix including all the vertebræ of the sample ($N = KI$). First, the within-group parameters are obtained after a particular decomposition of the matrix X. This decomposition consists in mean-centering all the sub-matrices X_i of size $K \times J$, where X_i is the partition of the matrix X belonging to the group i. Let $X_{c,i}$ be the resulting matrix. $X_{c,i}$ actually represents the within-patient variability. Therefore, each line of the matrix defines a vertebra. To obtain a global representation of the spine, matrices $X_{c,i}$ are transformed in a column-vector by concatenating the lines of $X_{c,i}$. Let X_c be the matrix built from the vertical concatenation of the resulting column-vectors. The matrix ϕ_W is composed of the eigenvectors of the covariance matrix related to X_c. Furthermore, the variance of the weight d_W, which limits the deformation of the within-group sub-model, is determined by the eigenvalues of the covariance matrix related to X_c.

In order to determine the between-group parameters of the model, let us subtract the overall mean of the matrix X. Let $\tilde{X}$ be this matrix. Let us note $\tilde{X}_i$, the partition of the matrix $\tilde{X}$ belonging to the group i. Moreover, let us consider the I vectors $\tilde{m}_i$, each of them representing the mean of the associated matrix $\tilde{X}_i$. In fact, these vectors characterize the between-patient differences. Therefore, we note M, the matrix resulting from the vertical concatenation of the vectors $\tilde{m}_i$. As a consequence, the matrix ϕ_B of the between-group principal components is built by the eigenvectors of the covariance matrix related to M. In addition, the deformation limits of the between-group sub-model are given by the eigenvalues of the covariance matrix related to M.

These concepts can naturally be extended to a greater number of hierarchical levels. Since we want to represent the variability between the vertebræ (*i.e. within-patient*), the extra levels are within-group terms. We can generalize the relation (2) in accordance with:

$$x_i = \bar{x} + \sum_{l=1}^{L-1} \phi_{W_l,i} d_{W_l} + \phi_B d_B, \tag{3}$$

where L is the number of levels and i_l is the index of the group to which the object belongs at the level l.

2.2   Reconstruction Algorithm

The principle of the reconstruction algorithm is to deform a 3D shape of the spine so that it matches the multiple views of the object to be reconstructed. In our case, two views are available: a posteroanterior (PA) and a lateral (LAT) radiograph. However, the final solution needs to be in conform to the model defined in section 2.1. In other words, the model constitutes a statistical *a priori* that the reconstruction algorithm has to take advantage of. A common metric to determine the degree of similarity between a shape and a shape model is the Mahalanobis distance.

The minimization of the Mahalanobis distance during the reconstruction process allows that the final shape fits the statistical distribution of the model. Another constraint needs to be met in the optimization problem. In order to match the 3D shape with the PA and LAT views, the Euclidean distance between the projection of a 3D point x_i and its theoretical location on the radiograph has to be minimized. Authors showed [2] that this distance can be computed given:

$$\left\| u_i^j - \tilde{u}_i^j \right\|_2 = \frac{1}{P_3^j \left(x_i\ 1 \right)^T} \left\| \begin{pmatrix} P_1^j - P_3^j u_{i,x}^j \\ P_2^j - P_3^j u_{i,y}^j \end{pmatrix} \begin{pmatrix} x_i \\ 1 \end{pmatrix} \right\|_2. \tag{4}$$

where P_i^j is the ith line of the matrix P^j and u_i^j is the projection of the point x_i. These authors also proposed to limit this error to a given constant e_{max} while optimizing for the point position. They formulated the problem as a second-order cone optimization program, just as it is demonstrated in [6]. Since second-order cone programming expects to operate with the norm of expressions, the Mahalanobis distance requires to be formulated with L, a Cholesky decomposition of Σ^{-1}, the inverse covariance matrix of the sample [2].

The second-order cone program is expressed by minimizing the Mahalanobis distance while constraining the solution to result in a projection error smaller than e_{max} using:

$$\begin{cases} \min & t \\ s.t. & \left\| L^T (x - \bar{x}) \right\|_2 \le t \\ & \left\| \begin{pmatrix} P_1^j - P_3^j u_{i,x}^j \\ P_2^j - P_3^j u_{i,y}^j \end{pmatrix} \begin{pmatrix} x_i \\ 1 \end{pmatrix} \right\|_2 \le e_{max} P_3^j \begin{pmatrix} x_i \\ 1 \end{pmatrix} \end{cases}. \tag{5}$$

To reduce the number of variables, it is possible to optimize MCA weights instead of point coordinates. To this end, let us define some specific notations. Let x_{ik} be a point k belonging to the vertebra x_i. Equation (3) allows to express:

$$x_{ik} = \bar{x}_k + \sum_{l=1}^{L-1} \phi_{W_l,ik} d_{W_l} + \phi_{B,k} d_B, \tag{6}$$

where $\bar{x}_k$, $\phi_{W_l,ik}$ and $\phi_{B,k}$ are respectively the lines of $\bar{x}$, $\phi_{W_l,i}$, and ϕ_B, associated to the point x_{ik}. Let us also consider $\sigma_{W_l}^2$ and σ_B^2, the variances associated to, respectively, the within-group and the between-group sub-models. Moreover, for simplicity of writing, let us define:

$$\psi(d_{W_l}, d_B)_k = \bar{x}_k + \sum_{l=1}^{L-1} \phi_{W_l,ik} d_{W_l} + \phi_{B,k} d_B. \tag{7}$$

Finally, the second-order cone programming optimization problem to match a multilevel statistical model with radiographic views is formulated as:

$$\begin{cases} \min \quad t \\ s.t. \quad \left(\sum_{l=1}^{L-1} \left\| diag(\frac{1}{\sigma_{W_l}}) d_{W_l} \right\|_2^2 + \left\| diag(\frac{1}{\sigma_B}) d_B \right\|_2^2 \right)^{\frac{1}{2}} \leq t \\ \quad \left\| \begin{pmatrix} P_1^j - P_3^j u_{k,x}^j \\ P_2^j - P_3^j u_{k,y}^j \end{pmatrix} \begin{pmatrix} \psi(d_{W_l}, d_B)_k \\ 1 \end{pmatrix} \right\|_2 \leq e_{max} P_3^j \begin{pmatrix} \psi(d_{W_l}, d_B)_k \\ 1 \end{pmatrix} \end{cases} . \tag{8}$$

3   Results

In order to process a reconstruction, two radiographs are presented to the user. He then has to point out some anatomical landmarks on the images to initiate the optimization of the problem (8). To validate our approach, we used a sample of 307 scoliotic patients for building the multilevel model. We considered 17 vertebræ of the spine: T1 to L5. Each vertebra is represented by 6 points of reference, *i.e.* the center of inferior and superior endplates, and the inferior and superior extremities of pedicles. 3D reconstructions based on a 2-level and a 3-level model have been performed for 25 post-operative patients whose spine was previously reconstructed following a reference method [1]. Let us note that tests were performed on an Intel Core 2 Duo 2.53 GHz.

We first present at Table 1 the decomposition of the total variability for a 2-level and a 3-level model. These values are computed in the same way as for the variance decomposition in ANOVA. Results of Table 1 show that the magnitude of the within-group and the between-group variability is sufficient to use a 2-level and a 3-level model in the reconstruction algorithm.

Mean RMS reconstruction error has been evaluated as a function of the number of points per radiograph (see Fig. 3). The parameter e_{max} was set to 8 pixels. The reconstruction of pedicles and plates are distinguished. One can remark that

Table 1. Decomposition of the total variability

Levels	Var. With. (%)		Var. Between (%)
2	88.83		11.17
3	23.12	65.71	11.17

plates reconstruction is better than pedicles reconstruction. This observation is actually similar in the case of reconstruction of the spine with no instrumentation [9]. Moreover, the reconstruction error decreases as the number of points per radiograph increases. However, this effect is reduced after approximatively 6 to 7 points per radiograph. Finally, Fig. 3 indicates that, both for plates and pedicles, the 3-level model converges more slowly than the 2-level model or the 1-level model. Adding extra levels in the model requires more constraints in the optimization problem. As a consequence, more control points per radiograph are needed. We have also compared the mean reconstruction error based on the 2-level model with a classical statistical model (with only one level) for 17 control points. A mean error of $2.12mm$ for endplates and $4.02mm$ for pedicles was obtained. For the 2-level model, these values are, respectively, $2.05mm$ and $3.70mm$. A paired *t-test* shows that these differences are significant (at level $\alpha = 0.05$). Actually, when a few control points are considered, the difference between the mean error of the 2-level model and the 1-level model is low. This difference increases with the number of control points. Furthermore, if we only consider instrumented vertebræ, the difference between the 2-level model and the classical model is increased. In this context, the 2-level model shows a mean error of $2.09mm$ for endplates and of $3.64mm$ for pedicles. The classical model is characterized by a mean error of $2.19mm$ for endplates and of $4.37mm$ for pedicles. This demonstrates that our approach based on a multilevel model is to be preferred in the case of 3D reconstruction of the post-operative patient spine.

Finally, we propose at Fig. 3 the execution times for a reconstruction based on a PCA model, a 2-level and a 3-level model. Since more constraints are considered
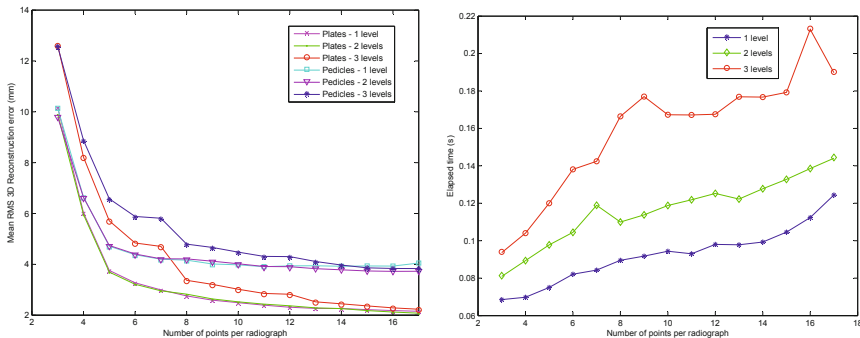
Fig. 3. Left: Evolution of the mean RMS 3D reconstruction error as a function of the number of points per radiograph. Right: Evolution of the elapsed time as a function of the number of points per radiograph

for the 3-level model, computing times are logically higher. Nevertheless, the results are an order of magnitude fast than most current methods, for example Humbert *et al.* [5] take about $4000ms$ to generate a reconstruction and Moura *et al.* [9] about $3000ms$. These results tend to show that our approach could be used interactively in the clinic.

4 Conclusion

In this manuscript, we proposed an algorithm to perform 3D reconstructions of the spine from bi-planar radiographs when surgical instrumentation is present. Our approach is based on a multilevel statistical model. Results showed that this model allows better reconstruction than classical models. The separation into several levels allows to deal with discontinuities characterizing the spine of post-operative patients. Since 3D spine reconstructions are obtained in real-time, preliminary results tend to show that our approach could be transposable to medical practice.

References

1. Aubin, C.E., Dansereau, J., Parent, F., Labelle, H., de Guise, J.A.: Morphometric evaluations of personalised 3D reconstructions and geometric models of the human spine. Medical & Biological Engineering & Computing 35(6), 611–618 (1997)
2. Boisvert, J., Moura, D.C.: Interactive 3D Reconstruction of the Spine from Radiographs Using a Statistical Shape Model and Second-Order Cone Programming. In: 33rd Annual International Conference of the IEEE EMBS, pp. 5726–5729 (2011)
3. de Bruijne, M., Lund, M.T., Tankó, L.B., Pettersen, P.C., Nielsen, M.: Quantitative vertebral morphometry using neighbor-conditional shape models. Medical Image Analysis 11(5), 503–512 (2007)
4. Cronbach, L.J., Webb, N.: Between-class and within-class effects in a reported aptitude treatment interaction: Reanalysis of a study by G. L. Anderson. Journal of Educational Psychology 67(6), 717–724 (1975)
5. Humbert, L., De Guise, J.A., Aubert, B., Godbout, B., Skalli, W.: 3D reconstruction of the spine from biplanar X-rays using parametric models based on transversal and longitudinal inferences. Medical Engineering & Physics 31(6), 681–687 (2009)
6. Kahl, F., Hartley, R.: Multiple-View Geometry Under the L_∞-Norm. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(9), 1603–1617 (2009)
7. Lecron, F., Boisvert, J., Benjelloun, M., Labelle, H., Mahmoudi, S.: Multilevel statistical shape models: a new framework for modeling hierarchical structures. In: IEEE International Symposium on Biomedical Imaging, pp. 1284–1287 (2012)
8. Mitulescu, A., Semaan, I., De Guise, J.A., Leborgne, P., Adamsbaum, C., Skalli, W.: Validation of the non-stereo corresponding points stereoradiographic 3D reconstruction technique. Med. Biol. Eng. Comput. 39(2), 152–158 (2001)
9. Moura, D.C., Boisvert, J., Barbosa, J.G., Labelle, H., Tavares, J.M.R.S.: Fast 3D reconstruction of the spine from biplanar radiographs using a deformable articulated model. Medical Engineering & Physics 33, 924–933 (2011)
10. Timmerman, M.E.: Multilevel component analysis. British Journal of Mathematical and Statistical Psychology 59(2), 301–320 (2006)

Probabilistic Segmentation of the Lumen from Intravascular Ultrasound Radio Frequency Data

E. Gerardo Mendizabal-Ruiz and Ioannis A. Kakadiaris

Computational Biomedicine Lab, Departments of Computer Science,
Electrical and Computer Engineering, and Biomedical Engineering,
University of Houston, Houston, TX

Abstract. Intravascular ultrasound (IVUS) is a catheter-based medical imaging technique that produces cross-sectional images of blood vessels. In this paper, we present a method for the segmentation of the luminal border using IVUS radio frequency (RF) data. Specifically, we parameterize the lumen contour using Fourier series. This contour is deformed by minimizing a cost function that is formulated using a probabilistic approach in which the *a priori* term is obtained using the prediction confidence of a Support Vector Machine classifier using features extracted from the RF signal. We evaluated the performance of our method by comparing our results with manual segmentations from two expert observers on 280 frames from eight 40 MHz IVUS sequences from rabbits and pigs. The performance was evaluated using the Dice similarity coefficient, coefficient of determination, and linear regressions of the lumen area for each frame. Our results indicate the feasibility of our method for the segmentation of the lumen from IVUS RF data.

1 Introduction

Intravascular ultrasound (IVUS) is a catheter-based imaging technique that provides high-resolution, cross-sectional images of the interior of blood vessels in real time. The IVUS system consists of a transducer which transmits pulses and receives a reflected radio frequency (RF) signal (i.e., A-line) at a discrete set of angles. These signals are then processed to reconstruct an image that is meaningful to the physicians (i.e., B-mode image). Accurate segmentation of IVUS images is important in order to assess vessel characteristics (e.g., the lumen and wall diameters), and for other applications (e.g., the study of mechanical properties of the vessel wall and the characteristics of the plaque, 3D reconstruction of the vessel). However, manual segmentation of IVUS data can be very expensive considering that one typical sequence may be composed of thousands of frames. Therefore, methods for automatic segmentation of IVUS data are needed.

Early IVUS systems operated at frequencies in the range of 10 to 20 MHz. At these frequencies, the blood presents a low acoustic impedance and therefore these systems produce IVUS images in which the lumen has low intensity, no texture, and a high contrast with respect to the vessel wall tissues. For this reason, many approaches for IVUS segmentation were based on the use of local properties of the image such as pixel intensity and gradient information. Modern IVUS systems operate at high frequencies

(i.e., 30 to 40 MHz) and produce images with high resolution. In these images, the lumen presents some texture due to speckle, and lower contrast with respect to the vessel wall tissues. For these images, edge information is not sufficient and therefore, later approaches incorporated prior knowledge using region and global information such as texture [1], gray level variances [2,3], statistical properties of the intensities [4], temporal information (3D segmentation) [5], and discrete wavelet decomposition [6]. Most recent approaches include the use of nonparametric probability densities with global measurements [7], multilevel discrete wavelet frame decomposition [8], discrete wavelet packet transform [9], machine learning classification methods [10], a combination of gray level probability density functions and the intensity gradient [11], linear-filtered gradient vector flow which drives the deformation of a balloon snake [12], and binary morphological object reconstruction [13].

A common characteristic of these methods is that the segmentation is performed using the reconstructed B-mode images. This poses a limitation considering that, apart from the frequency of operation, the appearance of the B-mode images depends on the reconstruction settings (e.g., time gain compensation (TGC), dynamic range compression, brightness, contrast and scaling) [14] (Figs. 1(a) and 1(b)). These settings are subjectively selected by the interventionist, and may change from one intervention to the next, or even during the same acquisition [15].

One solution to overcome these limitations is to perform the segmentation employing the raw IVUS RF signal since it is not affected by transformation parameters or visualization settings. RF-based approaches include methods for the characterization of different regions of interest such as plaque or blood (e.g., [16,17]).

In this work, we present a probabilistic method for the segmentation of the lumen from IVUS RF data. Our main contribution is a novel approach for the computation of the *a priori* terms necessary for a probabilistic segmentation by the analysis of the RF signal instead of the B-mode image. We evaluated the performance of our method by comparing our segmentation results with the manual segmentation from two expert observers on 280 frames from eight 40 MHz IVUS sequences from rabbits and pigs. Our results indicate the feasibility of our method for the segmentation of lumen.

The rest of the paper is organized as follows: Section 2 presents the steps of our segmentation method, Section 3 presents the results obtained with our method, and Sections 4 and 5 present our discussion and conclusion, respectively.

2 Methods

The proposed method is a significant improvement of our method presented in [18]. In that work, the segmentation of the lumen was achieved by the minimization of a probabilistic cost function formulated using a Bayesian approach:

$$U(\boldsymbol{C}) = \sum_r \sum_\theta P_l(r, \theta, \boldsymbol{C})[-log(v_l(r, \theta))] + [1 - P_l(r, \theta, \boldsymbol{C})][-log(v_n(r, \theta))] \,, \ (1)$$

where r and θ represent the polar coordinates of each pixel (i.e., radius and angle, respectively), P_l is the probability of each pixel to belong to the lumen class, and $v_l(r, \theta)$ and $v_n(r, \theta)$ represent to the *a priori* information (i.e., likelihood of the pixel to belong

to the lumen and non-lumen, respectively). The probability of a pixel to belong to lumen class is defined by its radial distance from the lumen curve $S(\theta, \boldsymbol{C})$ using a sigmoid function: $P_l(r, \theta, \boldsymbol{C}) = \left(1 + \exp^{-\lambda(S(\theta, \boldsymbol{C}) - r)}\right)^{-1}$.

Using this formulation, those pixels far above the contour will have a higher probability of belonging to the lumen, while the pixels far below the contour will have probability close to zero. For the pixels near the contour, depending on the value of λ, the probability of belonging to lumen will be close to 0.5.

The two major limitations of our previous method [18] are the number of parameters to be tuned for the parameterization of the curve which uses a mixture of Gaussians, and the simple method for estimating the *a priori* information models using gray-level histograms which limited the method for the segmentation of 20 MHz IVUS sequences exclusively. The proposed method has two significant differences with respect to that work: (i) we propose a parameterization which requires less parameters to be tuned and that suits better to the characteristics of the segmentation problem, and (ii) the computation of the *a priori* terms is based on the analysis of the RF signal.

The RF IVUS data corresponding to a cross section of a vessel consist of a set of one-dimensional signals. Each of these signals corresponds to the acoustic echoes acquired at a particular angle using a single-rotating transducer or a circular array of transducers. Therefore, the data can be analyzed using polar coordinates where the radius corresponds to the penetration of the ultrasound beam measured from the transducer (Fig. 1(d)). In this context, the lumen contour can be represented as a one-dimensional periodic curve. In this work, we have chosen to define the lumen curve using one-dimensional Fourier series instead of other parameterizations due to its simplicity and because this parameterization provides a periodic curve for which the smoothness can be controlled by the number of harmonics (i.e., number of coefficients N_k). The lumen contour is defined by:

$$S(\theta, \boldsymbol{C}) = \frac{a_0}{2} + \sum_{k=1}^{N_k} \left[a_k \cos\left(\frac{k2\pi\theta}{N_\theta}\right) + b_k \sin\left(\frac{k2\pi\theta}{N_\theta}\right) \right], \qquad (2)$$

where N_θ represents to the number of A-lines in a frame (i.e., width of the polar B-mode image), and $\boldsymbol{C} = [a_0, a_1, ..., a_{N_k-1}, b_1, ..., b_{N_k-1}]^T$ are the Fourier coefficients that control the shape of the curve.

For the computation of the *a priori* terms, we employ features extracted from the IVUS RF signal using our blood detection method proposed in [17]. In that work, the structures present in the vessel are modeled as a distribution of random positioned scatterers for which the acoustic power scattered in the direction of the ultrasound transducer is defined by the differential backscattering cross section (DBC) of the scatterer. The RF signal corresponding to each A-line is divided into N_P non overlapping partitions of constant width. The DBC of the scatterers that generates the signal of each partition is computed by minimizing the difference between the root mean square power of the real RF signal and the modeled RF signal computed with the model. In this work, we employ the DBC and the radial distance of the partition as features to detect the blood and non-blood regions using a machine learning method for classification.

In summary, our method consists of six steps: (i) divide the RF signal corresponding to each A-line of a frame into N_P non-overlapping partitions; (ii) compute the DBC

values for all the partitions using the method described in [17]; (iii) construct a feature vector $v_p \in R^2$ for each partition p by concatenating the DBC computed for that partition and its radial distance from the transducer; (iv) classify each partition as blood or non-blood according to the prediction of a classifier; (v) use the confidence of the classification results as the *a-priori* term in the Bayesian optimization function; and (vi) minimize the cost function that guides the position and shape of the lumen contour using a line-search method.

In this work, we chose to use a Support Vector Machines (SVM) classifier in order to take advantage of its ability of mapping the data to a higher dimensional space on which the problem becomes linearly separable. The training and deployment steps for the SVM model are described next.

Training: In the training step, the user to provides examples of lumen and non-lumen regions by manual annotations of these regions on the first frame B-mode reconstruction of the sequence to be segmented (Figs. 1(b) and 1(c)). Next, the DBC of the scatterers that generate the signal of each partition corresponding to the regions provided as examples is computed. A two-dimensional feature vector $v_p \in R^2$ is constructed for each partition. A training set is defined by associating each vector with its corresponding class and then used to generate an SVM model that is trained using a Gaussian kernel. The optimal parameters c and γ for the SVM model are computed using grid search and 5-fold cross validation.

Deployment: For each of the frames to be segmented, the DBC of the scatterers for each partition of every A-line is computed, and the feature vector is constructed in the same way as in the training step. The classification of each partition is performed using the SVM model giving as a result a class and an associated confidence P_s (Fig. 1(e)). This classification confidence is then used as the likelihood for blood and non-blood as $v_l(r, \theta) = P_s(r, \theta)$ and $v_n(r, \theta) = 1 - P_s(r, \theta)$, respectively. Finally, the segmentation is performed by the minimization of the cost function of Eq. (1) using the steepest descent method.

3   Results

We use the proposed method to perform segmentation on selected frames from eight 40 MHz IVUS sequences, four were acquired from rabbits and the other four from pigs. For each sequence, we selected five groups of seven consecutive frames from different periods of time for a total of 35 frames per sequence. The parameters used for the computation of the DBC are $\sigma = 5.3e^{-8}$ for the width of the envelope of the impulse function, $\mu = 0.08276$ *dB/mm* corresponding to the attenuation coefficient of blood, $c = 1,540 \times 10^3$ *mm/s* corresponding to the speed of sound in biological tissue, $\triangle P = 0.05$ *mm* for the size of partition, $D = 400$ scatterers/mm^{-2} for the scatterer density, and $\delta = 3$ for the cardinality of neighbors according to [17]. The number of coefficients used for the contour parameterization was $N_k = 5$. The parameter for the probabilistic segmentation was experimentally set to $\lambda = 0.4$. The initial point for the
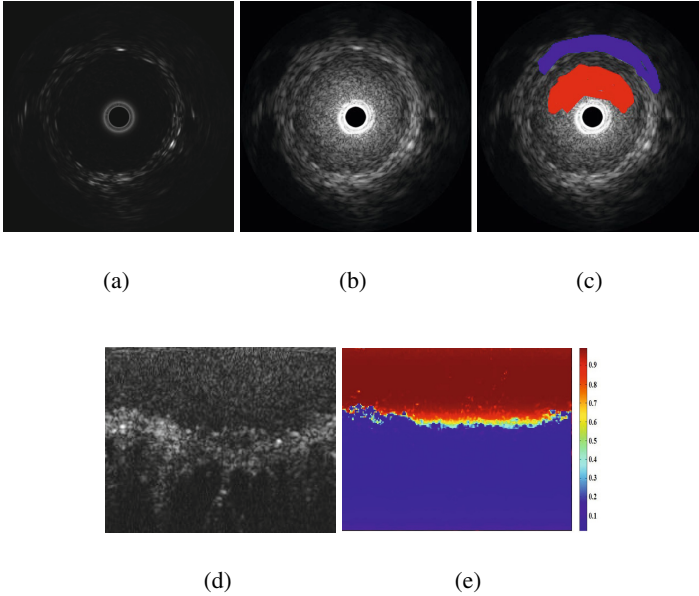
Fig. 1. Example depicting (a) the first frame of a sequence in Cartesian B-mode representation using logarithmic dynamic range compression, (b) the same frame using linear dynamic range compression with a compression factor of 0.4, (c) the corresponding user annotation for blood and non-blood (red and blue, respectively), (d) the polar B-mode representation of the frame, and (e) the probability of each partition to belong to blood according to the SVM model in polar representation (color figure).

steepest descent optimization was empirically set to $a_0 = 1$, $a_i = 0.1$ and $b_i = 0.1$ $\forall\, i > 0$. The method was implemented in MATLAB and the average time for training the SVM for each sequence was $12\,s$ and the segmentation time per frame was $5.4\,s$ on an Intel i7 at 2.67 GHz 12GB RAM.

We evaluated the performance of our method by comparing our segmentation results (A) with the manual segmentation from two expert observers (O_1 and O_2). We computed the Dice similarity coefficient for the comparison of segmentations of (O_1 vs. O_2), (O_1 vs. A), and (O_2 vs. A) for each frame. Figure 2 depicts the results of this analysis for each sequence and for the total number of frames. Note that our segmentation results are similar to the segmentation provided by the two expert observers. In addition, we evaluated the agreement between the areas defined by each segmentation by computing the coefficient of determination and using linear regressions (Fig. 3). Although there exists a good agreement between the areas, a tendency of our method to under-segment the areas can be observed. Finally, Figure 4 depicts examples of the segmentation results along with the manual segmentations from the two expert observers.
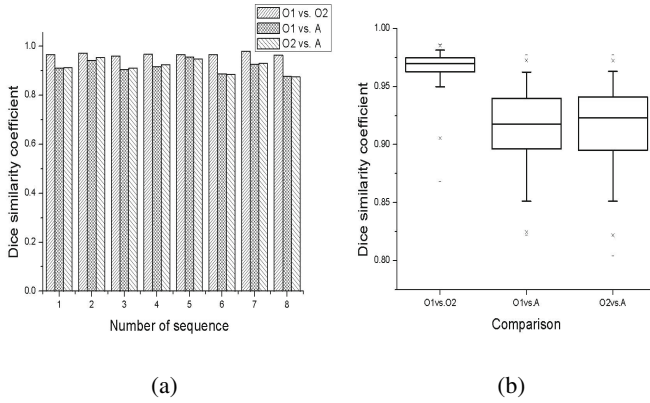
Fig. 2. Mean Dice similarity coefficient for (a) each sequence, and (b) the total number of frames
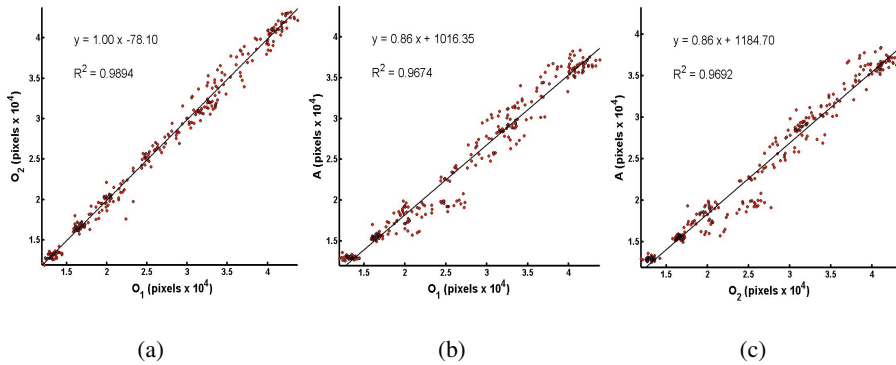
Fig. 3. Linear regression and coefficient of determination for the comparisons of (a) (O_1 vs. O_2), (b) (O_1 vs. A), and (c) (O_2 vs. A)
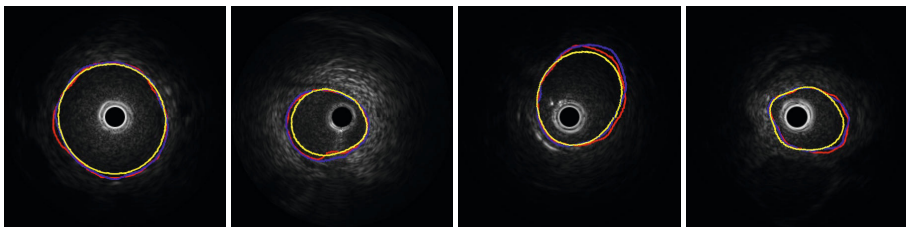
Fig. 4. Examples of segmentation results (yellow) and their qualitative comparison with manual segmentations

4 Discussion

The main advantage of our method when compared with other existing IVUS segmentation methods is that our method is not affected by the B-mode reconstruction parameters since it is based on the use of the RF signal. While there may be a concern regarding the availability of such data, note that recent IVUS systems available in the clinic are capable of providing the RF signal. The selected dataset included frames with moderate guidewire artifacts and did not include frames with strong guidewire artifacts, plaque, stents and side branches as this is subject for future work. In this work, we do not evaluate the accuracy of the classification method in the classification of a frame since this data is not used as a direct result of the segmentation method but instead as a likelihood of a pixel to belong to a certain class. The final class assigned to each pixel is determined by the segmentation curve. Real-time segmentation may be achieved by an implementation in a lower level language and by segmenting frames in parallel. Additionally, the proposed method may be fully automated by creating an SVM model using samples from a large dataset of sequences instead of training for each sequence. The main limitation of our method is the requirement of user initialization. Additionally, our method may provide an incorrect segmentation in the presence of side branches for which expert knowledge is required. Improvements on the classification performance, the use of other classification methods, different parameterizations of the lumen curve, exploration of additional features, including additional classes apart from blood and non-blood (e.g., guidewire, shadow, plaque.), segmentation of the media/adventitia interface, the use of temporal information (i.e., 3D-approach), and comparison with other existing methods are subjects of future research.

5 Conclusion

We have presented a novel approach for the segmentation of the luminal border using IVUS RF data. Our results indicate the feasibility of our method for the segmentation of lumen in real time applications.

References

1. Mojsilovic, A., Popovic, M., Amodaj, N., Babic, R., Ostojic, M.: Automatic segmentation of intravascular ultrasound images: A texture-based approach. Annals of Biomedical Engineering 25(6), 1059–1071 (1997)
2. Haas, C., Ermert, H., Holt, S., Grewe, P., Machraoui, A., Barmeyer, J.: Segmentation of 3D intravascular ultrasonic images based on a random field model. Ultrasound in Medicine and Biology 26(2), 297–306 (2000)
3. Luo, Z., Wang, Y., Wang, W.: Estimating coronary artery lumen area with optimization-based contour detection. IEEE Transactions on Medical Imaging 22, 564–566 (2003)
4. Brusseau, E., de Korte, C.: Fully automatic luminal contour segmentation in intracoronary ultrasound imaging - a statistical approach. IEEE Transactions on Medical Imaging 23(5), 554–566 (2004)
5. Cardinal, M., Meunier, J., Soulez, G., Maurice, R., Therasse, E., Cloutier, G.: Intravascular ultrasound image segmentation: A three-dimensional fast-marching method based on gray level distributions. IEEE Transactions on Medical Imaging 25(5), 590–601 (2006)

6. dos Santos Filho, E., Yoshizawa, M., Tanaka, A., Saijo, Y.: A study on intravascular ultrasound image processing. Record of Electrical and Communication Engineering Conversazione 74(2), 30–33 (2006)
7. Unal, G., Bucher, S., Carlier, S., Slabaugh, G., Fang, T., Tanaka, K.: Shape-driven segmentation of the arterial wall in intravascular ultrasound images. IEEE Transactions on Information Technology in Biomedicine 12(3), 335–347 (2008)
8. Papadogiorgaki, M., Mezaris, V., Chatzizisis, Y., Giannoglou, G., Kompatsiaris, I.: Image analysis techniques for automated IVUS contour detection. Ultrasound in Medicine and Biology 34(9), 1482–1498 (2008)
9. Katouzian, A., Baseri, B., Konofagou, E., Laine, A.: Automatic detection of blood versus non-blood regions on intravascular ultrasound (IVUS) images using wavelet packet signatures. In: Proc. SPIE Medical Imaging 2008: Ultrasonic Imaging and Signal Processing, San Diego, CA, February 16-21 (2008)
10. Ciompi, F., Pujol, O., Fernández-Nofrerías, E., Mauri, J., Radeva, P.: ECOC Random Fields for Lumen Segmentation in Radial Artery IVUS Sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 869–876. Springer, Heidelberg (2009)
11. Cardinal, M., Soulez, G., Tardif, J., Meunier, J., Cloutier, G.: Fast-marching segmentation of three-dimensional intravascular ultrasound images: A pre-and post-intervention study. Medical Physics 37(7), 3633–3647 (2010)
12. Zhu, X., Zhang, P., Shao, J., Cheng, Y., Zhang, Y., Bai, J.: A snake-based method for segmentation of intravascular ultrasound images and its in vivo validation. Ultrasonics 51(2), 181–189 (2011)
13. Moraes, M., Furuie, S.: Automatic coronary wall segmentation in intravascular ultrasound images using binary morphological reconstruction. Ultrasound in Medicine and Biology 37, 1486–1499 (2011)
14. Hiro, T., Leung, C., Russo, R., Karimi, H., Farvid, A., Tobis, J.: Variability of a three-layered appearance in intravascular ultrasound coronary images: A comparison of morphometric measurements with four intravascular ultrasound systems. American Journal of Cardiac Imaging 10(4), 219–227 (1996)
15. Mintz, G., Nissen, S., Anderson, W., Bailey, S., Erbel, R., Fitzgerald, P., Pinto, F., Rosenfield, K., Siegel, R., Tuzcu, E., Yock, P.: American College of Cardiology clinical expert consensus document on standards for acquisition, measurement and reporting of intravascular ultrasound studies (IVUS). Journal of the American College of Cardiology 37(5), 1478–1492 (2001)
16. Nair, A., Margolis, M., Kuban, B., Vince, D.: Automated coronary plaque characterisation with intravascular ultrasound backscatter: Ex vivo validation. Eurointervention 3(1), 113–120 (2007)
17. Mendizabal-Ruiz, E., Biros, G., Kakadiaris, I.A.: Towards Extra-Luminal Blood Detection from Intravascular Ultrasound Radio Frequency Data. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 396–403. Springer, Heidelberg (2011)
18. Mendizabal-Ruiz, E., Rivera, M., Kakadiaris, I.: A probabilistic segmentation method for the identification of luminal borders in intravascular ultrasound images. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Anchorage, AK, June 24-26, pp. 1–8 (2008)

Precise Segmentation of Multiple Organs in CT Volumes Using Learning-Based Approach and Information Theory

Chao Lu[1,3], Yefeng Zheng[1], Neil Birkbeck[1], Jingdan Zhang[1], Timo Kohlberger[1], Christian Tietjen[2], Thomas Boettger[2], James S. Duncan[3], and S. Kevin Zhou[1]

[1] Image Analytics and Informatics, Siemens Corporate Research, Princeton, New Jersey, USA
[2] Computed Tomography, Siemens Healthcare, Forchheim, Germany
[3] School of Engineering & Applied Science, Yale University, New Haven, Connecticut, USA

Abstract. In this paper, we present a novel method by incorporating information theory into the learning-based approach for automatic and accurate pelvic organ segmentation (including the prostate, bladder and rectum). We target 3D CT volumes that are generated using different scanning protocols (e.g., contrast and non-contrast, with and without implant in the prostate, various resolution and position), and the volumes come from largely diverse sources (e.g., diseased in different organs). Three key ingredients are combined to solve this challenging segmentation problem. First, marginal space learning (MSL) is applied to efficiently and effectively localize the multiple organs in the largely diverse CT volumes. Second, learning techniques, steerable features, are applied for robust boundary detection. This enables handling of highly heterogeneous texture pattern. Third, a novel information theoretic scheme is incorporated into the boundary inference process. The incorporation of the Jensen-Shannon divergence further drives the mesh to the best fit of the image, thus improves the segmentation performance. The proposed approach is tested on a challenging dataset containing 188 volumes from diverse sources. Our approach not only produces excellent segmentation accuracy, but also runs about eighty times faster than previous state-of-the-art solutions. The proposed method can be applied to CT images to provide visual guidance to physicians during the computer-aided diagnosis, treatment planning and image-guided radiotherapy to treat cancers in pelvic region.

1 Introduction

Automated segmentation of medical images has increasingly become a valuable tool for medical image analysis. Pelvic region analysis plays an important role in the medical diagnosis and treatment planning for prostate cancer and bladder cancer. Segmentation of prostate as well as bladder and rectum from a three dimensional computed tomography (CT) volume often serves as the first step in image-based radiotherapy studies and continuously attracts research attention [1,2,3,4,5,6,7]. Though intensive research has been performed, accurate segmentation of 3D soft tissue structures in pelvic region is still a challenging problem, due to the large variations in organ shapes and in the texture pattern inside and along organ boundaries.

(a) 0.81 × 0.81 × 5 mm   (b) 1.0 × 1.0 × 1.0 mm   (c) 0.98 × 0.98 × 3 mm   (d) 0.98 × 0.98 × 5.4 mm

Fig. 1. Examples of CT volumes in our dataset. Note the variance in the image size, position of prostate / bladder / rectum, and volume dimensions. (c) also presents a challenging case that the patient is undergoing brachytherapy, with implanted metal seeds in the prostate.

Unlike in previous work, the CT volumes in our study are scanned under largely diverse protocols. Specifically, there are several diversities in our dataset: 1) The volume dimension and organ position in the volume substantially vary; 2) Some volumes are enhanced by contrast agent while some are not; 3) Some volumes come from the patients undergoing brachytherapy with implanted metal seeds in the prostate; 4) The inter-slice resolution changes from 0.8 mm to 5.0 mm. Some examples are shown in Figure 1. These diversities increase the variability of the key organs in both shape and texture patterns, and therefore render the problem more challenging. Furthermore, the automatic detection (or localization) is made more difficult especially because of 1). Albeit important, automatic object detection is largely ignored in previous work.

In this paper, we propose an automatic segmentation approach that addresses all the challenges mentioned above. Figure 2 illustrates the flowchart of the proposed method. There are three major contributions in our system. First, the system uses a marginal space learning strategy, which efficiently and effectively solves the organ detection problem. Second, a learning-based boundary localization technique is utilized. By using this technique, the system not only achieves accurate boundary responses, but also becomes reliable to the heterogeneous texture patterns. Third, an information theory based module is incorporated into the framework and the Jensen-Shannon divergence-based uncertainty measures are used for further improvement.
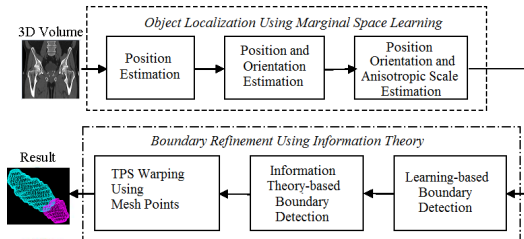
Fig. 2. Diagram for the proposed method

2   Organ Detection and Mesh Initialization - Hierarchical Detection Network

We represent the shape of an organ by a dense collection of mesh points on its boundary, therefore, by a closed triangle mesh. In our work, the prostate is constructed with 602 points, while the bladder and rectum are represented using 1202 and 962 points, respectively.

We initialize the multi-organ segmentation from an existing learning-based detection framework. In the first stage, we performed a hierarchical bounding box detection system based on Marginal Space Learning [8] in connection with Probabilistic Boosting Trees [9]. The latter yields bounding box estimates for the prostate, the bladder, and the rectum.

After automatic object localization, we align the mean shape with the estimated pose onto the 3D volume, which is used as the initialization for mesh detection. An illustration of mesh initialization for bladder is presented in Figure 3.
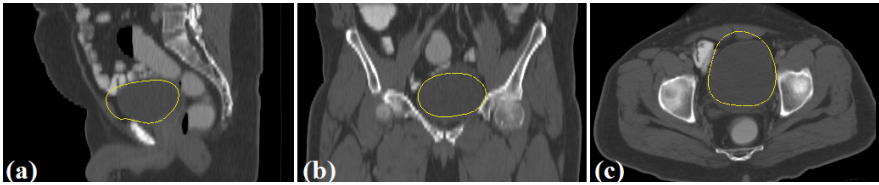
Fig. 3. Mesh initialization after pose detection for the bladder. (a)(b)(c) show the sagittal, coronal and axial view, respectively.

3   Boundary Refinement

After initialization, we then perform the boundary refinement to accurately fit the object boundary. Active shape models (ASM) are widely used to deform an initial estimate of a deformable shape under the guidance of the image evidence and the shape prior. However, the generic boundary detector in the original ASM does not work in our application due to the complicated background and weak edges. There are two key components in our refinement method: probabilistic boundary response, and information theoretic measure (Jensen-Shannon divergence).

3.1   Learning-Based Boundary Detection

A key issue in boundary refinement is boundary localization, which usually involves locally searching around current shape boundaries. Specifically, for a current boundary point $p \in P$, a candidate point set Q_p is formed by including points along the normal direction at p and within some distance. Then the point in Q_p with the maximum boundary probability (response) is used to replace p, i.e.,

$$p \leftarrow \arg\max_{q \in Q_p} Pr(boundary|q, vol). \tag{1}$$

Previous work usually approximates the boundary response by simple checking gradients or intensity distribution along surface normals. However, the information gathered this way is not enough for our task because the texture pattern in the pelvic region has a large variability. When dealing with data from different scanning protocols, the variability is even larger.

In order to attack this problem, we decide to learn $Pr(boundary|q, vol)$ from the training set. In this paper, we train a boundary detector using our efficient steerable features and PBT, similarly to [8]. The same approach is used to train a boundary detector using the PBT [9] and steerable features. According to our experiments, the learning-based approach provides a robust boundary inference and it is reliable to the heterogeneous texture patterns, however, there is still large potential for improving the precision.

3.2   Information Theory Incorporation

In soft-tissue segmentation, there is often no clear boundary information. The intensity gradient at the interface of adjacent tissues is also low. To overcome the difficulty, besides the learning-based method, we also incorporate the information theory based class-uncertainty information into our boundary refinement work, using Jensen-Shannon divergence measure.

Jensen-Shannon (JS) divergence, first introduced in [10], serves as a measure of cohesion between multiple probability distributions. It has been used as a dissimilarity measure for image registration with promising results. It has several desirable properties, to name a few, 1) The square root of JS-divergence is a metric [11]; 2) JS-divergence relates to other information theoretic functionals, such as the relative entropy or the Kullback divergence, and hence it shares their mathematical properties as well as their intuitive appeal; 3) The compared distributions using the JS-divergence can be weighted, which allows one to take into account the different sizes of the point set samples from which the probability distributions are computed; 4) the JS-divergence is bounded by 0 and 1.

In this paper, the JS-divergence is the main measure that we incorporate into the boundary refinement framework. Given a priori knowledge of intensity probability distributions of each object, the JS-divergence can be computed for each intensity value, which yields the uncertainty level of the classification. Suppose the image is divided into m objects, denoted by $o_i, i \in [0, 1, ..., m-1]$. For any candidate point q in Q_p with image intensity $I(q)$, the posteriori probability $p(o_i|I)$ is obtained from the training set using Bayes rules:

$$P_i(I) = p(o_i|I) = \frac{p(I|o_i) \cdot p(o_i)}{p(I)}, \tag{2}$$

where $p(I|o_i)$, $p(o_i)$, and $p(I)$ are learned from the training set.

The JS-divergence of each intensity I is defined by:

$$JS(I) = JS(P_1(I), P_2(I), ..., P_i(I)) = H(\Sigma \pi_i P_i(I)) - \Sigma \pi_i H(P_i(I))$$
$$= H(\Sigma \pi_i p(o_i|I)) + \Sigma(\pi_i \cdot p(o_i|I) \log(p(o_i|I))), \tag{3}$$

where $\pi = \{\pi_1, \pi_2, ..., \pi_n | \pi_i > 0, \Sigma\pi_i = 1\}$ are the weights of the probability distributions P_i, and $H(\mathbf{x})$ is the Shannon entropy, i.e., $H(x) = \int -p(x)\log p(x)dx$. The two terms on the right hand side are the entropy of $\mathbf{P} = \Sigma\pi_i P_i(I)$ (the π-convex combination of the P_is) and the same convex combination of the respective entropies.

In [12], the authors defined a simple uncertainty measure, which is a specific case of our general Jensen-Shannon divergence, but the mathematical properties demonstrated in [12] still hold for JS divergence. Intensities with high JS-divergence measures tend to appear close to object boundaries rather than inside homogeneous regions. Combined with the gradient information, the JS-divergence measure is expected to provide an enhanced on-surface cost at the locations absent a clear boundary. Let $\nabla(q)$ denote the gradient magnitude at a candidate point q, our information theoretic measure has the following form:

$$
\begin{aligned}
p &\leftarrow \arg\max_{q \in Q_p} C(I(q), vol) \\
&= \arg\max_{q \in Q_p} \left[\nabla(q) + \alpha JS(q)\right],
\end{aligned}
\tag{4}
$$

where α is a constant parameter. Here we set $\alpha = 0.5$ according to the experiments on the training set. For simplicity, we choose $\pi_i = p(o_i)$, and in this paper, the volume is divided into four parts: prostate, bladder, rectum and the background. It is straightforward to see that $\Sigma\pi_i = 1$.

After performing the control point adjustment described above, the mesh points fit the boundary well, but the contour may not be smooth. Then, we project the deformed shape onto a shape subspace. In all our experiments, to determine the dimension of the subspace, we demand it to capture 95% variations. As shown in previous work [8], the statistical shape model is very effective to enforce the prior shape constraint.

4   Experiments

4.1   Dataset

Our database contains 188 3D CT volumes, each with three annotated groundtruth dense meshes for prostate, bladder and rectum, respectively. As mentioned in the introduction, the dataset is very challenging in that the volumes come from largely diverse sources. This heterogeneity causes a large variation in both shape deformation and texture patterns of the organs. Moreover, diagnosis of different diseases often request different contrast agent to be injected into the patient, or no contrast at all.

4.2   Evaluation

We conducted the evaluation using a four-fold cross validation. Errors are measured using the average symmetric surface distance. The average running time for one volume, including all steps, is around 1.06 seconds for the prostate, 1.51 seconds for the bladder and 2.53 seconds for the rectum (on an Intel 8-Core 2.00 GHz processor, with 16G RAM).

Figure 4 illustrates a difficult case that the patient is undergoing brachytherapy in prostate. The learning-based approach cannot lock on the edge of prostate while the proposed method produces satisfying result. The comparison of bladder segmentation is also shown in Figure 4.
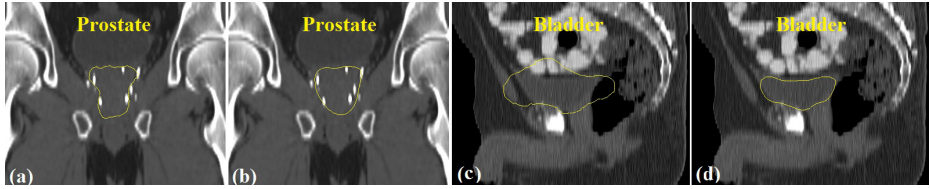
Fig. 4. Segmentation of the prostate and bladder, using (a)(c) the learning-based approach only, and (b)(d) the proposed method

To give a thorough evaluation, in this paper, we use four-fold cross validation to evaluate our algorithm. Specifically, the whole dataset (188 cases) is divided to four sets, each containing 47 volumes. Each time one set is chosen as the testing set and the rest as the training set (for both the shape model and boundary classifiers). This is done for four times and the average performance is reported.

The evaluation results for cross validation are summarized in Table 1. The mean error measured in the average symmetric surface distance is 2.37 mm for the prostate, and the median is 2.15 mm. For the bladder, the mean and median error are 2.81 mm and 2.24 mm, respectively. The rectum achieves an accuracy of 4.23 mm and 4.09 mm for the mean and median, respectively.

Table 1. Cross Validation for the Proposed Method (mm)

	Method	Mean	STD	Median	Minimum	Maximum	80th percentile
Prostate	Learning-based [8]	3.53	2.23	3.00	1.52	13.06	5.01
	Proposed Method	2.37	0.89	2.15	1.10	5.58	2.96
Bladder	Learning-based [8]	3.89	1.89	3.70	0.87	10.39	5.53
	Proposed Method	2.81	1.86	2.24	0.82	8.76	3.38
Rectum	Learning-based [8]	6.44	1.71	6.19	3.65	12.46	7.86
	Proposed Method	4.23	1.46	4.09	1.81	8.19	5.24

It is difficult to directly compare different segmentation approaches due to the use of different datasets as well as different annotations. However, it is worth summarizing the previous experiments to comprehend the status of the study. Table 2 provides a brief summary of recent works on automatic pelvic region segmentation with reported average surface distances along with datasets used. The table shows clearly that our precision is among the the best reported. Compared to the method [1] with similar reported precisions, our approach have two apparent strengths. First, the dataset used in our experiments has more diversities than previous tested datasets. Second, our method is more efficient, i.e. runs more than eighty times faster.

Table 2. Comparison with Recent Work on Prostate Segmentation

Method	Mean Error	Speed	Data set	Auto?	Interslice Res.	Implanted?	Organs
Freedman *et al.* [5]	3.91 *mm*	NA	48	Yes	NA	No	P, R
Rousson *et al.* [3]	4 to 5 *mm*	12 s	16	Semi	3 *mm*	No	P, B
Feng *et al.* [1]	2.47 *mm*	96 s	24	Semi	3 *mm*	No	P
Proposed Method	2.37 *mm*	1.06 s	188	Yes	Varies	Mixed	P, B, R

Some typical segmentation results are shown in Figure 5 from three orthogonal views, with segmented prostate in yellow, bladder in blue and rectum in green.
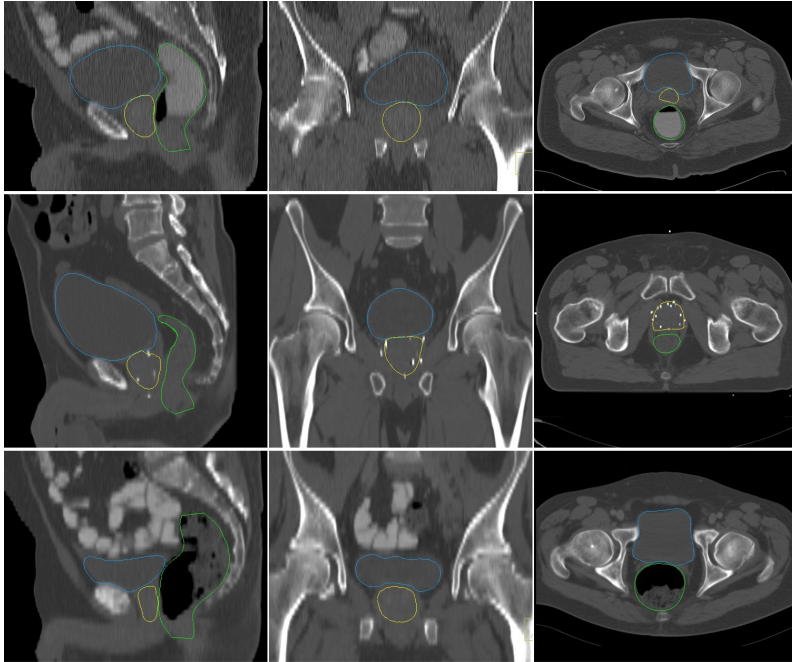
Fig. 5. Examples of pelvic-region multiple organ segmentation, with yellow for the prostate, blue for the bladder, and green for the rectum. From left to right: sagittal, coronal, and axial slices.

5   Conclusion

In this paper, we proposed a novel approach that incorporates information theory with the learning-based approach, for automatic pelvic-region multiple organ segmentation from a 3D CT volume. We target on general data from patients with different diseased organs and scanned under different protocols. Despite the challenges, our approach demonstrates robust performance in accuracy and runs more efficiently than state-of-the-art solutions. Using the techniques described above, it is possible for the physicians to efficiently delineate the key pelvic organs for diagnosis and treatment planning, and also precisely guide the interventional devices toward the organs during image-guided radiotherapy.

References

1. Feng, Q., Foskey, M., Chen, W., Shen, D.: Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. Medical Physics 37(8), 4121–4132 (2010) 462, 467, 468

2. Lu, C., Chelikani, S., Papademetris, X., Knisely, J.P., Milosevic, M.F., Chen, Z., Jaffray, D.A., Staib, L.H., Duncan, J.S.: An integrated approach to segmentation and nonrigid registration for application in image-guided pelvic radiotherapy. Medical Image Analysis 15(5), 772–785 (2011) 462

3. Rousson, M., Khamene, A., Diallo, M., Celi, J.C., Sauer, F.: Constrained Surface Evolutions for Prostate and Bladder Segmentation in CT Images. In: Liu, Y., Jiang, T., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 251–260. Springer, Heidelberg (2005) 462, 468

4. Lu, C., Duncan, J.S.: A coupled segmentation and registration framework for medical image analysis using robust point matching and active shape model. In: 2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), pp. 129–136 (January 2012) 462

5. Freedman, D., Radke, R., Zhang, T., Jeong, Y., Lovelock, D., Chen, G.: Model-based segmentation of medical imagery by matching distributions. IEEE Transactions on Medical Imaging 24(3), 281–292 (2005) 462, 468

6. Lu, C., Chelikani, S., Jaffray, D.A., Milosevic, M.F., Staib, L.H., Duncan, J.S.: Simultaneous nonrigid registration, segmentation, and tumor detection in MRI guided cervical cancer radiation therapy. IEEE Transactions on Medical Imaging 31(6), 1213–1227 (2012) 462

7. Seifert, S., Barbu, A., Zhou, S.K., Suehling, M., Cavallaro, A., Comaniciu, D.: Hierarchical parsing and semantic navigation of full body ct data. In: SPIE Medical Imaging (February 2009) 462

8. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. on Med. Imaging 27(11), 1668–1681 (2008) 464, 465, 466, 467

9. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: ICCV 2005, vol. 2, pp. 1589–1596 (October 2005) 464, 465

10. Lin, J.: Divergence measures based on the Shannon entropy. IEEE Trans. Infor. Theory 37, 145–151 (1991) 465

11. Endres, D., Schindelin, J.: A new metric for probability distributions. IEEE Trans. Inf. Theory 49, 1858–1860 (2003) 465

12. Saha, P., Udupa, J.: Optimum threshold selection using class uncertainty and region homogeneity. IEEE Trans. Pattern Anal. Machine Intell. 12, 689–706 (2001) 466

A Study on Graphical Model Structure for Representing Statistical Shape Model of Point Distribution Model

Yoshihide Sawada and Hidekata Hontani

Nagoya Institute of Technology, Aichi, Japan

Abstract. In this article, the authors demonstrate that you can improve the performance of the registration of a point distribution model (PDM) by accurately estimating the structure of an undirected graphical model that represents the statistical shape model (SSM) of a target surface. Many existing methods for constructing SSMs determine the structure of the graphical model without analyzing the conditional dependencies among the points in PDM, though an edge in the PDM should link two nodes if and only if they are conditionally dependent. In this study, the authors employed four popular methods for estimating the structure of graphical model and obtained four different SSMs from an identical set of training surfaces. The registration performances of the SSMs were experimentally compared, and the results showed that the graphical lasso, which could estimate more accurate structure of the graphical model by avoiding the overfitting to the training data, outperformed the other methods.

Keywords: Registration, Point Distribution Model, Graphical Model, Graphical lasso.

1 Introduction

It is an important step for automated image analysis to register a surface model of a target organ with given 3D medical images [1]. A point distribution model (PDM) is widely used for representing the surface model [2,3,4]. A PDM represents the surface using a set of points distributed on the surface, and is registered to images by estimating the coordinates of each of the points. The statistical shape model (SSM) of PDM plays an important role in this registration [1].

The SSM of a target organ is constructed from a set of training surfaces. A set of corresponding points is generated on each of the surfaces, and the SSM is constructed based on the distributions of the corresponding points. One of the most popular SSMs is the active shape model (ASM) [1,5], which represents the statistics of the target surface by using a linear low-dimensional subspace obtained by applying PCA to the corresponding points. When ASM is employed, one can register the model by iteratively projecting target points to the subspace. This projection is an ML estimation of the location of the model surface. Another popular SSM for PDM is an undirected graphical model [3,4]. In the graphical model, each node represents the coordinates of each of the points in the PDM. When the undirected graphical model is employed, one can register the model by inferring the coordinates on the graphical model by means of, e.g., MCMC [3]. The strength of this latter approach is that not only the coordinates of each

point but also its posterior marginal distribution can be obtained [4]. In other words, the confidence of the estimated location of each point can also be evaluated. In this study, we employ the undirected graphical model for representing the SSM of PDM.

The aim of this study was to investigate the relationship between the structure of the graphical model and the performance of the registration. In the graphical model, the nodes represent the variables and the edges represent the conditional dependencies of the variables: an edge links two nodes if and only if the corresponding two variables are conditionally dependent [6]. Few methods for registering PDM, though, determine the structure of the graphical model based on the analysis of the conditional dependencies. For example, many existing methods employing the undirected graphical model determine the structure based on the geodesic distances [4,7]: Two nodes are linked by an edge if the geodesic distance between the corresponding two points are enough short [4]. It is sure that the locations of such two points are conditionally dependent together, but is *not* sure that only such two points are conditionally dependent. The ASM, on the other hand, implicitly represents the SSM using a complete graph: All nodes are linked together. It is known that the exact structure of the graphical model can be determined based on the precision matrix of the variables if they obey the Gaussian: The precision matrix is the inverse of the covariance matrix, and the non-zero off-diagonal entries of the precision matrix correspond to the edges [8]. The subspace representation in ASM is constructed from the covariance matrix, and its inverse is dense in general. This means that all points on a target surface are implicitly assumed to be conditionally dependent each other. It looks too strong assumption. The performance of the registration is improved when the SSM represents the statistics of the target surface more accurately.

There have been proposed some approaches for estimating the structure of a graphical model based on a set of training data [8,9]. The authors employ a graphical lasso [8], which estimates the precision matrix of a given set of training data with sparse regularization. Applying the graphical lasso, you can explicitly estimate the sparse structure of the graphical model, which represents the SSM of the PDM.

Employing different approaches for the estimation of the graphical model structure, one obtains different SSMs from an identical set of training data. In this study, four different approaches are employed for determining the structure of the graphical model: (1) the graphical lasso, (2) the ASM, (3) the canonical correlation-based approach, and (4) the geodesic distance-based one. In the canonical correlation-based approach, two nodes are linked by an edge if the canonical correlation between the corresponding two points is strong enough [3]. Obtained four different SSMs of a target organ, each model was registered to an identical set of test images and the accuracy of the resultant registration was compared. Then, we experimentally found that the graphical lasso constructed the best model structure among them.

2   Surface Registration Based on Inference on Graphical Model

In this article, we use a non-rigid surface registration method proposed in [3,4]. The outline of the method is described in this section.

2.1 Surface Model Construction

A surface is represented with a set of N points $\{P_j | j = 1, 2, \cdots, N\}$. Let $\boldsymbol{x}_j$ denote the 3D coordinates of P_j. A statistical model of the points is constructed based on a set of M training images. Let $\mathcal{I} = \{I^i | i = 1, 2, \cdots, M\}$ denote the set of the training images. These images are normalized based on the body shape in advance. Let S^i denote the surfaces of the target organ extracted from I^i manually.

For constructing the statistical model, a set of N corresponding points, $\{P_j | j = 1, 2, \cdots, N\}$, are distributed on S^i. For distributing the corresponding points, the entropy-based particle system [10] is used. Let I^i_j denote the local appearance in a $L \times L \times L$ cube centered at P^i_j. Given a set of $\{P^i_j, I^i_j | i = 1, 2, \cdots, M, j = 1, 2, \cdots, N\}$, the method estimates the following three distributions: $p(\boldsymbol{x}_j)$, $p(I_j|\boldsymbol{x}_j)$, and $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$. $p(\boldsymbol{x}_j)$ denotes the prior distribution of P_j, $p(I_j|\boldsymbol{x}_j)$ denotes the conditional probability of the local appearance and is used for obtaining the likelihood distribution of P_j, and $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$ denotes the probability distribution of the relative location. Using these probabilities, the simultaneous probability distribution is represented as follows:

$$p(\{\boldsymbol{x}_j\}, \{I_j\}) = \frac{1}{Z} \prod_j p(\boldsymbol{x}_j) p(I_j|\boldsymbol{x}_j) \prod_{e_{j,k} \in \mathcal{E}} p(\boldsymbol{x}_j, \boldsymbol{x}_k), \tag{1}$$

where $e_{j,k}$ denotes an edge in the graphical model that links P_i and P_j, and $\mathcal{E}$ denotes a set of the edges. In this article, $p(\boldsymbol{x}_j)$ is represented by a pillbox distribution and $p(I_j|\boldsymbol{x}_j)$ and $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$ are represented by the Gaussian. The estimation of these probability distributions is straightforward when the training set, $\{P^i_j, I^i_j\}$, is given. The only problem is to determine the set of the edges, $\mathcal{E}$.

As mentioned in the previous section, four approaches for determining $\mathcal{E}$ are employed for the investigation. Each of them is described in the followings.

Graphical Lasso (GL). It often happens that the empirical precision matrix overfits to the training data especially when the number of the data is small. The graphical lasso avoids this overfitting by means of the L1-norm regularization. Let $\boldsymbol{X}^i = [\boldsymbol{x}^{i\top}_1, \boldsymbol{x}^{i\top}_2, \cdots, \boldsymbol{x}^{i\top}_N]^\top$. Given a set of $\mathcal{X} \equiv \{\boldsymbol{X}^i | i = 1, 2, \cdots, M\}$, the graphical lasso estimates the sparse precision matrix, $\hat{\boldsymbol{\Theta}}$, of $\mathcal{X}$ by maximizing the following criterion [8].

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \{\log \det \boldsymbol{\Theta} - \text{tr}(\boldsymbol{S}\boldsymbol{\Theta}) - \rho ||\boldsymbol{\Theta}||_1\}, \tag{2}$$

where $||.||_1$ denotes the L1-norm, and $\boldsymbol{S}$ denotes the empirical covariance matrix of $\mathcal{X}$. Once $\hat{\boldsymbol{\Theta}}$ is obtained, $e_{j,k}$ is added to $\mathcal{E}$ if and only if the (j, k) entry of $\hat{\boldsymbol{\Theta}}$ is non-zero. The parameters of $p(\boldsymbol{x}_j, \boldsymbol{x}_k)$ can also be determined based on $\hat{\boldsymbol{\Theta}}$. The value of ρ will be determined based on the performance of the registration.

Active Shape Model (ASM). Let $\boldsymbol{\Theta}' \equiv \boldsymbol{S}^{-1}$. In this study, all eigen vectors of $\boldsymbol{S}$ are used for constructing the subspace representation. Then, $e_{j,k}$ is added to $\mathcal{E}$ if and only if (j, k) entry of $\boldsymbol{\Theta}'$ is non-zero. In general, all off-diagonal entries in $\boldsymbol{\Theta}'$ are non-zero when the number of training surfaces, M, is small, because the precision matrix overfits to the training data. This means that $\mathcal{E}$ represents a set of the edges of the complete graph.

(A)   (B)
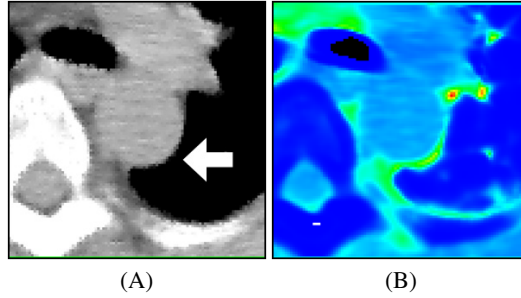
Fig. 1. An example of given images, I_j (A) and the likelihood distribution of P_j computed by $p(I_j|\boldsymbol{x}_j)$ (B). The white arrow in (A) points at the location of P_j.

Canonical Correlation-Based Approach (CC). Let $c_{j,k}$ denote the canonical correlation [3] between $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ computed from $\{(\boldsymbol{x}_j^i, \boldsymbol{x}_k^i)|i = 1, 2, \cdots, M\}$. In this approach, $e_{j,k}$ is added to $\mathcal{E}$ if and only if $c_{j,k} > T$, where T is a threshold, which will be determined based on the performance of the registration. It should be noted that P_j and P_k are not always conditionally dependent, even if $c_{j,k}$ is large.

Geodesic Distance Based Approach (GD). Let $d_{j,k}$ denote the average geodesic distance between P_j^i and P_k^i. In this appraoch [4], $e_{j,k}$ is added to $\mathcal{E}$ if and only if $d_{j,k} < D$, where D is a threshold.

2.2   Registration of Surface Model

Given a new X-CT image, the method registers the surface model to the image by inferring the marginal posterior distribution $p(\boldsymbol{x}_j|I)$. For this inference, we can use the Gibbs sampling (GS) [3]. It should be noted that $p(I_j|\boldsymbol{x}_j)$ is represented by the Gaussian, but the likelihood distirbution of $\boldsymbol{x}_j$ does not obey the Gaussian, as shown in Fig. 1. This is why we apply the nonparametric GS for the inference. The registration algorithm is as follows [3]:

1. Normalize the body shape in a given image [4].
2. Compute $p(\boldsymbol{x}_j)p(I_j|\boldsymbol{x}_j)$.
3. Apply nonparametric GS for inferring the marginal posterior distribution, $p(\boldsymbol{x}_j|I)$, based on the simultaneous distribution in (1).

The MAP estimate of $\boldsymbol{x}_j$ is obtained as $\boldsymbol{x}_j^* \equiv \arg\max_{\boldsymbol{x}_j} p(\boldsymbol{x}_j|I)$, and the expectation of $\boldsymbol{x}_j$ is computed as $E[\boldsymbol{x}_j] = \int_{\boldsymbol{x}_j} p(\boldsymbol{x}_j|I)d\boldsymbol{x}_j$. In this study, the error of the registration at P_j is defined as $|E[\boldsymbol{x}_j] - Q_j|$, where Q_j is the closest point to P_j on the answer surface extracted manually for the error evaluation [3].

3   Experimental Results

We selected the arch of aorta for the target organ. Given a set of $M = 12$ non-contrast enhanced CT images, of which resolution is 0.98mm $\times$ 0.98mm $\times$ 4.25mm,
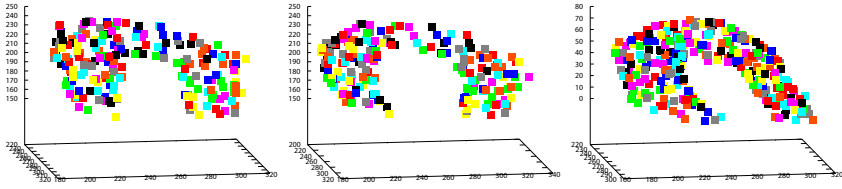
Fig. 2. Examples of corresponding points, P_j^i. Each color indicates the number j.

(A)   (B)   (C)
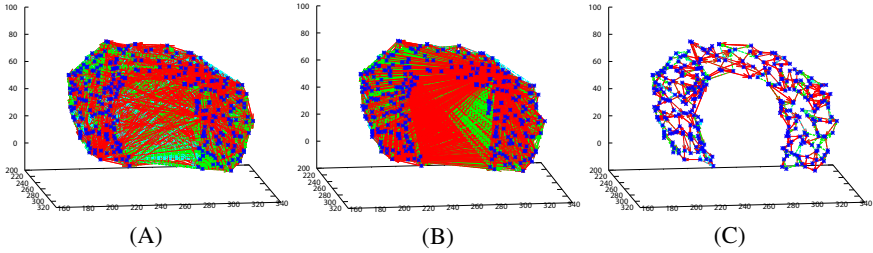
Fig. 3. Examples of the estimated structures of the graphical model. The graphs were estimated by the graphical lasso (A), the canonical correlation-based approach (B), and the geodesic distance-based one (C), respectively.

we constructed four different graphical models of the aorta by using the four different approaches described in sec.2.1. The registration accuracy of them was evaluated by means of the leave-one-out cross validation. Each model had $N = 300$ points for the representation. Figure 2 shows examples of the points.

Figure 3 shows three graphical models estimated by (A) the graphical lasso (GL), (B) the canonical correlation-based approach (CC), and (C) the geodesic distance-based one (GD), respectively. The model obtained by the ASM is omitted because the graph has too many edges to indicate clearly. As shown in the figure, many pairs of distant two points were linked by edges in the models obtained by the GL and by the CC. Some characteristics of the models are shown in Table 1. As indicated in this table, the model constructed by GL had smaller number of edges than that by CC. The maximum degree

Table 1. Characteristics of the structures of the models obtained by GL (graphical lasso), ASM (active shape model), CC (canonical correlation-based approach) and GD (geodesic distance based approach).

	# of edges	# of degrees		
		minimum	average	maximum
GL	7834	11	52	94
ASM	44850	299	299	299
CC	10029	25	66	298
GD	545	3	3	7

Table 2. The registration errors (mm) obtained by each of the models. The minimum, the average, and the maximum values of the errors are indicated.

	minimum	average	maximum
GL	1.67	3.26	5.40
ASM	1.88	3.48	5.27
CC	1.76	3.46	5.43
GD	2.05	3.77	6.37
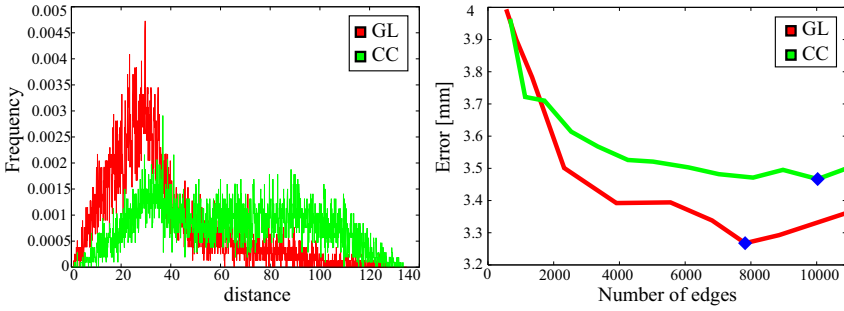
Fig. 4. (A) The histograms of the geodesic distance between two points linked by an edge in the graphical model. Red: graphical lasso. Green: Canonical correlation-based approach. (B) The change of the registration error (mm) with respect to the number of edges in the graphical model.

(A)   (B)   (C)
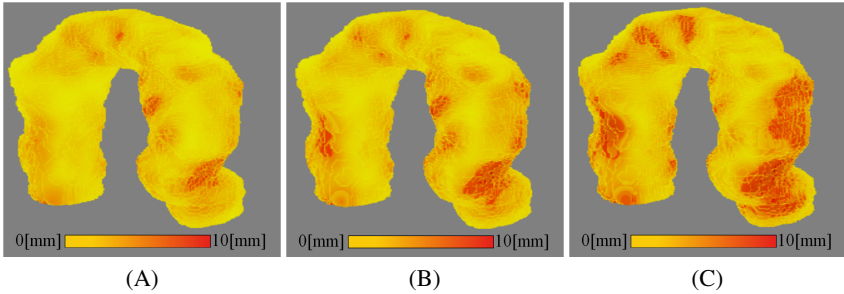
Fig. 5. Examples of the distributions of the registration errors. Each was obtained by registering the model constructed (A) by the graphical lasso, (B) by the ASM, and (C) by the canonical correlation-based approach.

in the model made by CC was 298, which was almost same with that in the ASM (299). This means that there was at least one node that was connected with almost all other nodes. The maximum degree was relatively small in the model made by the GL.

The difference between the two models constructed by the graphical lasso and by the canonical correlation analysis is demonstrated in Fig. 4 (A). The x-axis shows the geodesic length between two points linked by an edge, and the y-axis shows the number of the edges. The graphical model constructed by the graphical lasso had more short edges than that by the canonical correlation. The reason of this difference is as follows. The deformation of the aorta made by the normalization of the body shape was approximately symmetry with respect to the median line of the aorta [4]. This deformation generated strong correlations between two distant points located at opposite sides of the aorta, hence these two distant points were more easily linked by edges.

As mentioned in sec. 2.1, the graphical lasso and the canonical correlation-based method have parameters that control the sparseness of the graphical model. We investigated the change of the registration accuracy with respect to the parameters. Figure 4 (B) shows the result. The x-axis in the graph shows the number of edges in the model and the y-axis shows the registration error (mm) evaluated by the leave-one-out cross
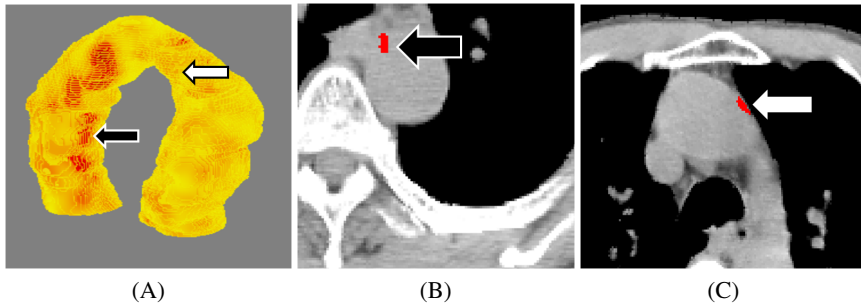
Fig. 6. (A) An example of the distribution of the variance σ_j^i, which denotes the standard deviation of the posterior marginal distribution along the normal of the surface. The region pointed by the black arrow has larger values of σ_j^i and that pointed by the white arrow has smaller values of σ_j^i. (B) The slice image that contains the point indicated by the black arrow in (A). No edge of the aorta can be observed. (C) The slice image that contains the point indicated by the white arrow in (A). The wall of the aorta can be clearly observed.

validation. As shown in the graph, the model constructed by the graphical lasso was superior to the model constructed by the canonical correlation analysis. We determined the values of the parameters based on this experiment, and selected the model for each that demonstrated the best performance. The selected models are indicated by a blue dots in the graph.

Table 2 shows the comparison of the four models with respect to the registration error. As shown in the table, the model constructed by the graphical lasso was the best among them. We applied the Wilcoxson rank-sum test and the null hypothesis was rejected ($p < 0.05$): In this case, we can surely decide that the graphical lasso outperformed other methods. It should be reminded that the graphical lasso estimates the structure of the graphical model based on the conditional dependencies among the variables and that the overfitting of the precision matrix is avoided by means of the L1-norm regularization. The authors believe that the superiority of the graphical lasso came from its ability to analyze the statistics of the training surfaces more accurately. Some examples of the distributions of the registration error are shown in Fig. 5. The red color density indicates the magnitude of the error. As shown in this figure, the distributions of the error were not uniform.

As mentioned earlier, the confidence of the registration can also be estimated based on the variance of the marginal posterior distribution along the normal direction of the surface. Let the variance be denoted by σ_j^i. Figure 6 (A) shows an example of the distribution of σ_j^i. The black arrow points at unconfident point at which σ_j^i is larger, and the white one points at more confident point where σ_j^i is smaller. Figure 6(B) shows a slice image containing the point indicated by the black arrow in (A). As shown in the image, the edge was not observed because of the contact with neighboring organs, and the method successfully judged that the registration at the location was unconfident. On the other hand, the point indicated by the white arrow in Fig. 6(A) was located on the high-contrasted edge in the image as shown in Fig. 6(C), and the method automatically judged that the registration result was confident.

4 Summary

An undirected graphical model is widely used for representing the SSM of a PDM, and its structure represents the conditional dependencies among the variables. In this study, the authors applied four different methods to a given set of the training data for estimating the structure of the graphical model, and obtained four different SSMs from an identical set of the training data. Then, these four SSMs were compared with respect to the accuracy of the registration and the experimental results showed that the graphical lasso outperformed the other methods. The graphical lasso determines the structure by estimating the precision matrix with the L1-norm regularization in order to avoid its overfitting to the training data. When the SSM of the PDM is represented by an undirected graphical model, you can improve the registration performance by determining the structure of the graphical model based on the statistical analysis of the conditional dependencies among the variables. In this study, the value of the parameter that controls the sparseness of the precision matrix was determined based on the registration accuracy, and this process was very time consuming. The future works include to develop an effective method for determining the value of the parameter.

References

1. Heimann, T., Meinzer, H.P.: Statistical shape models for 3D medical image segmentation: A review. Medical Image Analysis 13, 543–563 (2009)
2. Zhang, S., Zhan, Y., Dewan, M., Huang, J., Metaxas, D.N., Zhou, X.S.: Deformable Segmentation via Sparse Shape Representation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 451–458. Springer, Heidelberg (2011)
3. Sawada, Y., Hontani, H.: A Comparison Study of Inferences on Graphical Model for Registering Surface Model to 3D Image. In: Suzuki, K., Wang, F., Shen, D., Yan, P. (eds.) MLMI 2011. LNCS, vol. 7009, pp. 257–264. Springer, Heidelberg (2011)
4. Hontani, H., Watanabe, W.: Point-Based Non-Rigid Surface Registration with Accuracy Estimation. In: Computer Vision and Pattern Recognition, pp. 446–452 (2010)
5. Cootes, T.F., Taylor, C.J., Graham, J.: Active Shape Models-Their Training and Application. Computer Vision and Image Understanding 61, 38–59 (1995)
6. Allassoniere, S., Jolivet, P., Giraud, C.: Detecting Long Distance Conditional Correlations Between Anatomical Regions Using Gaussian Graphical Models. In: 3rd MICCAI Workshop on Mathematical Foundations of Computational Anatomy, pp. 111–122 (2011)
7. Zhang, P., Adeshina, S.A., Cootes, T.F.: Automatic Learning Sparse Correspondences for Initialising Groupwise Registration. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 635–642. Springer, Heidelberg (2010)
8. Friedman, J., Hastie, T., Tibshirari, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9, 432–441 (2008)
9. Bickel, P.J., Levina, E.: Covariance regularization by thresholding. The Annals of Statistics 36, 2577–2604 (2008)
10. Cates, J.E., Fletcher, P.T., Styner, M.A., Shenton, M.E., Whitaker, R.T.: Shape Modeling and Analysis with Entropy-Based Particle Systems. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 333–345. Springer, Heidelberg (2007)

Quality Metric for Parasternal Long Axis
B-Mode Echocardiograms

Sri-Kaushik Pavani[1], Navneeth Subramanian[1], Mithun Das Gupta[1],
Pavan Annangi[1], Satish C. Govind[2], and Brian Young[3]

[1] John F. Welch Technology Center, GE Global Research, Bangalore, India
[2] Bhagwan Mahavir Jain Heart Centre, Bangalore, India
[3] GE Healthcare-PCS, Milwaukee, WI, USA

Abstract. This paper presents a method for automatically estimating
the quality of Parasternal Long AXis (PLAX) B-mode echocardiograms.
The purpose of the algorithm is to provide live feedback to the user on
the quality of the acquired image. The proposed approach uses Gen-
eralized Hough Transform to compare the structures derived from the
incoming image to a representative atlas, thereby providing a quality
metric (PQM). On 133 PLAX images from 35 patients, we show: 1)
PQM has high correlation with manual ratings from an expert echocar-
diographer 2) PQM has high correlation with contrast-to-noise ratio, a
traditional indicator of image quality 3) on images with high PQM, error
in automatic septal wall thickness measurement is low, and *vice versa*.

1 Introduction

Standard views in transthoracic echo are well established. But, quality of images
and correct tomographic planes for accurate clinical interpretation and measure-
ments are dependent on operator skills. Algorithms that can automatically detect
quality of ultrasound images have tremendous potential to 1) standardize imag-
ing 2) reduce scan time for users by providing real-time feedback, and 3) provide
automatic mechanism to reject data of poor quality.

Assessment of image quality prior to complex post-processing is common in
the field of human identification using biometric data [1]. Our contribution ex-
tends this philosophy to medical ultrasound images. Note that our work is dif-
ferent from image quality testing of imaging systems [2] in that we propose an
approach to compute image quality during live imaging rather than on phan-
toms.

We propose an algorithm to automatically determine the quality of Paraster-
nal Long AXis (PLAX) B-mode echocardiograms. At the right scan plane, and
with optimal instrument settings, the long axis of the left ventricle is oriented
horizontally in a standard PLAX view (See Fig. 1(a)). The posterior wall, the
pericardium and the septum are approximately parallel to each other. Any devi-
ation from this is a result of an incorrect acquisition or sub-optimal instrument
settings. For example, the poor quality image shown in Fig. 1(b) could be due
to sub-optimal instrument settings such as gain, dynamic range, or time-gain
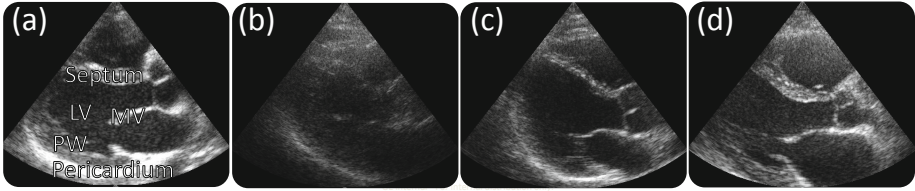
Fig. 1. Examples of good (a) and bad (b,c,d) quality Parasternal Long Axis (PLAX) images. Key: LV - left ventricle, MV - mitral valve, PW - posterior wall.

Original image   Contrast-enhanced   Segmented image   GHT
image
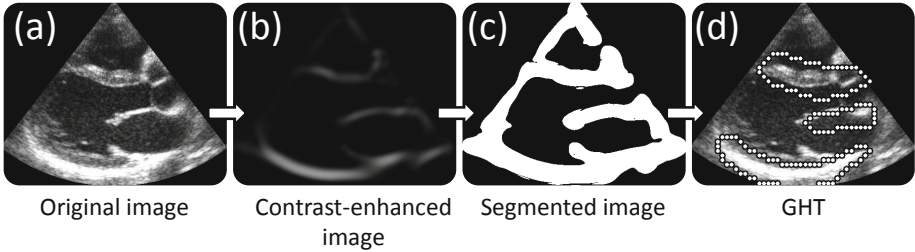
Fig. 2. Flow chart illustrating the proposed methodology to compute the quality metric

compensation. The non-parallel septum and pericardium in Fig. 1(c) tells us that the imaging plane did not pass through the center of left ventricle. In Fig. 1(d), the pericardium is missing, thereby, complicating the measurement of the thickness of posterior wall and diagnosis of pericardial effusion.

The intuition behind our algorithm is to check for the presence of three prominent tube-like structures corresponding to septum, mitral valve and pericardium. As visualized in Fig. 2, the proposed approach comprises of the following three steps. 1) A pre-processing step that enhances the contrast of thick tube-like structures. 2) A global thresholding step, that outputs a binary image which includes the three structures of interest: septum, mitral valve and pericardium. 3) A search algorithm based on Generalized Hough Transform (GHT) [3] [4], that best matches a pre-defined atlas with the binary image. The GHT outputs an accumulator image, whose maximum value is output as the PLAX Quality Metric (PQM).

2 Methodology

In the following, we present an algorithm that searches for septum, mitral valve, and the pericardium, and outputs PQM, which is indicative of how prominently these structures appear in the image.

Contrast Enhancement and Segmentation: As noted by Boukerroui *et al.* [5], segmentation algorithms based only on global information such as thresholding techniques, intensity dependent clustering and edge detection schemes

give poor results on ultrasound images. The reason being, the echo amplitude varies according to the orientation of the reflected structure and the image settings such as gain, dynamic range, or time-gain compensation. This is commonly referred to as the intensity inhomogeneity problem.

To mitigate intensity inhomogeneity, we filter the input echo image using Frangi vesselness filter [6] to obtain an intermediate image as shown in Fig. 2(b). The filtered image has less intensity inhomogeneity, and we have observed that global thresholding schemes are more successful on this intermediate representation.

Frangi's vessel enhancement filter is based on the eigen analysis of the Hessian matrix of image intensity at each pixel location of the image. The eigen analysis of the Hessian directly gives the direction of smallest curvature (along the tubular structure). The mutual magnitude of eigenvalues is indicative of the shape of the underlying object. Frangi *et al.* noted that if a pixel were to lie on a tubular structure, then, one of the eigenvalues has much higher magnitude than the other. If the pixel was from a background region, then both eigenvalues will have low magnitudes.

Let λ_1 and λ_2 denote the two eigenvalues the Hessian $\mathcal{H}_{o,s}$, computed at scale s at pixel x_o. Frangi *et al.* propose the following equation to obtain a vesselness image ($\mathcal{V}$).

$$\mathcal{V}_o = \begin{cases} 0 & \text{if } \lambda_2 > 0 \\ \exp\left(-\frac{\mathcal{R}_B^2}{2\beta^2}\right)\left(1 - \exp\left(-\frac{\mathcal{S}^2}{2c^2}\right)\right) & \text{otherwise} \end{cases}$$

Here $\mathcal{R}_B = \lambda_1/\lambda_2$, $\mathcal{S} = \sqrt{\lambda_1^2 + \lambda_2^2}$, β and c are constants which we set to 1 and 0.5, respectively. We experimentally observed that by thresholding $\mathcal{V}$ at 0.02, one obtains a binary image as illustrated in Fig. 2(c).

Commonly, depth parameter of the echocardiogram is set such that the pericardium is at the bottom of the image for PLAX images. We resized all PLAX images to 50×50 pixels, and we noted that the thickness of pericardium, mitral value, septum does not vary much. Therefore, we chose to detect vessels at one scale ($s = 6$ pixels). Although vascular enhancement at multi-scales might produce better images, we chose a single scale to reduce the processing time.

Note that the underlying theory behind this preprocessing step is not specific to vessels. A popular extension of Frangi's technique by Antiga [7] is capable of enhancing contrast of blob-like and plate-like structures.

Generalized Hough Transform (GHT): It is essentially a method to detect the presence of an arbitrary object (described with its model/atlas) in a binary image [3] [4]. In our case, the atlas is illustrated in Fig. 3(b), and the image to be searched for is the binary image obtained after thresholding the Frangi vesselness image (Fig. 2(c)).

The GHT uses a lookup table that encodes the relationship between the atlas and the Hough parameters. This lookup table is called the R-Table, and the Hough parameters, r, and α are computed during the training phase. See Algorithm 1 for details. The parameter r encodes the distance between a pixel and

---

Algorithm 1. GHT-based training

Input: Binary training image (atlas) of size $m \times n$: $\mathcal{I}$
Input: Reference point $\{(x_c, y_c) \mid 1 < x_c < m, 1 < y_c < n\}$
Output: R-table: $\mathcal{R}$
foreach $\mathbf{x} = \{(x, y) \mid \mathcal{I}(\mathbf{x}) = 1\}$ **do**
>   Compute following parameters:
> $r = \sqrt{(x_c - x)^2 + (y_c - y)^2}$
> $\alpha = \text{atan2}(y_c - y, x_c - x)$.
>   gradient direction ϕ
> $\phi = \text{arctangent}\left(\dfrac{\partial \mathcal{I}_t / \partial y}{\partial \mathcal{I}_t / \partial x}\right)$
> $\mathcal{R}(\phi) \leftarrow (r, \alpha)$

end

---

the reference point (which is chosen arbitrarily within the image). The parameter α encodes the angle, in radians, between the positive x-axis and the vector spanning a pixel and the reference point. During the matching phase, as shown in Algorithm 2, one attempts to find the most probable location of the atlas on a test image. The accumulator $\mathcal{A}$ is a 2D array that holds the votes received for the most probable location of the reference point. The pixel with the maximum intensity in $\mathcal{A}$ is the most probable location of the reference point. The maximum value of $\mathcal{A}$ is output as PLAX Quality Metric (PQM).

Although correlation based techniques [8] are a viable alternative to GHT-based matching, the choice of GHT was based on the interest in speeding up the search procedure. If the atlas and the segmented image have n pixels, then the correlation based procedure, in spatial domain, has the complexity O(n^2). In contrast, the GHT-based procedure has the complexity O(m^2) ($m << n$), where m is the number of white pixels in atlas and the segmented image.

PLAX Atlas: As mentioned above, GHT matching procedure (Algorithm 2) uses an atlas that defines the structures of interest in PLAX images. The atlas was generated via the following steps.

1. Manual segmentation of the regions of interest in PLAX images. An example is shown in Fig. 3(a).
2. Average representation of the structures of interest was obtained using Shape-Based Averaging (SBA) algorithm [9].

Several independent binary images result from the manual segmentation process, which must be somehow combined into a single final segmentation. Majority voting is the generally used rule to fuse the segmentations, but better methods, such as SBA, have been proposed. SBA consists of averaging Euclidean distance maps computed for all structures for each candidate manual segmentation. The method was shown to keep structure regularity and contiguity better than majority voting. From 89 manually segmented frames, we obtain a PLAX atlas as illustrated in Fig. 3(b).

Algorithm 2. GHT-based matching

Input: Binary test image of size $k \times l$: $\mathcal{I}_t$
Input: R-table (output of Algorithm 1): $\mathcal{R}$
Output: PLAX quality metric: PQM
$\mathcal{A}(x_c, y_c) \leftarrow 0 \colon \forall x_c \in \{1 \ldots k\}, \forall y_c \in \{1 \ldots l\}$
foreach $\mathbf{x} = \{(x, y) \mid \mathcal{I}_t(x, y) = 1\}$ **do**
    Compute gradient direction ϕ
    $\phi = \text{arctangent}\left(\dfrac{\partial \mathcal{I}_t / \partial y}{\partial \mathcal{I}_t / \partial x}\right)$
    Using ϕ, retrieve corresponding (r, α) values from $\mathcal{R}$
    foreach (r, α) **do**
        Compute candidate reference points:
        $x_c = x + r\cos(\alpha)$
        $y_c = y + r\sin(\alpha)$
        Increment accumulator (voting):
        $\mathcal{A}(x_c, y_c) = \mathcal{A}(x_c, y_c) + 1$
    end
end
$\text{PQM} = \max\limits_{x_c \in [1 \ldots k], y_c \in [1 \ldots l]} \mathcal{A}$
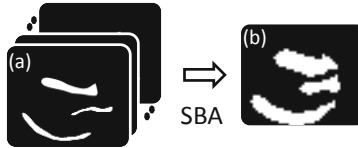
Fig. 3. A total of 89 manually segmented images (a) belonging to distinct patients were used to obtain the PLAX atlas (b)

3   Results

A total of 35 subjects with varied clinical background, normal chamber dimensions and normal systolic function underwent routine echocardiography (commercially available Vivid 7, GE) with electrocardiogram gating. The patient data used in our validation included normal and hypertrophic patients. The data was acquired by both echo-cardiologist and an echo-technician at a clinical site. PLAX images with 3 cardiac cycles were analyzed by an expert sonographer for grading image quality.

Comparison with Manual Ratings: The purpose of this experiment was to verify whether PQM correlates with an expert echocardiographist's ratings. The expert manually rated the quality of septum, mitral valve and pericardium on 133 PLAX echocardiograms belonging to 35 patients. Each of these components was given a rating between 0 and 3, with 0 signifying poor visibility and 3 signifying good visibility. The final manual score for an image was obtained as the average
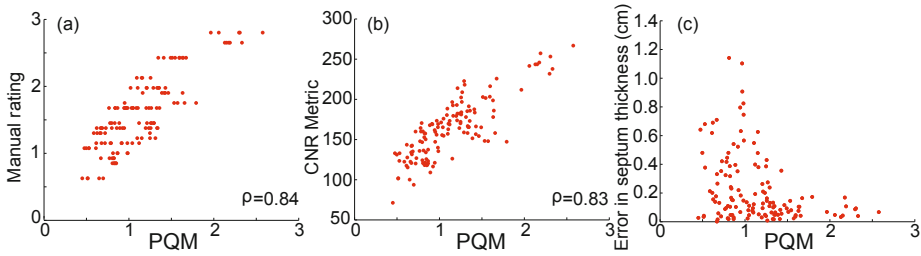
Fig. 4. Comparison with Contrast-to-Noise ratio. Scatter plot between Contrast-to-Noise (CNR) ratio and PQM. Scatter plot between error in septum thickness measurement and PQM.

of the individual scores. Thus, the manual score for an image varies between 0 and 3. To facilitate comparison with automatic ratings, we applied min-max normalization to the automatic ratings such that the value of PQM is between 0 and 3. Fig. 4(a) shows a plot between manual rating and PQM. As observed, PQM seems to correlate well (Pearson's correlation coefficient, $\rho = 0.84$) with an echocardiographist's manual ratings.

Comparison with Contrast-to-Noise Ratio: Contrast-to-noise ratio (CNR) has been traditionally used in medical imaging community to quantify the quality of acquired images. This metric removes the subjectivity factor from the manual ratings. CNR measures the ability to distinguish between an object of interest and its surroundings. The CNR can be defined as: $\text{CNR} = \frac{|\mu_o - \mu_s|}{\sigma}$. Here, μ_o and μ_s are the mean pixel intensities in the object and surrounding regions, respectively. The quantity σ is the standard deviation of the intrinsic noise of the imaging system.

Using manual segmentations of septum, mitral valve and pericardium (See Fig. 3(a) for an illustration), we obtained masks over the regions of interest, and thus, we could compute μ_o for each image. The immediate surrounding areas of the mask were used to compute μ_s. The noise parameter, σ, can be effectively ignored for our experiments because all the images were acquired using the same ultrasound machine.

On 133 PLAX echocardiograms, the CNR metric and the PQM were computed. The results are shown as a scatter plot between the two quantities in Fig. 4(b). We observed a Pearson's correlation coefficient, ρ, of 0.83.

Comparison with Error in Septum Thickness Measurement: Subramanian *et al.* [10] proposed a snakes-based approach that automatically measures septal wall thickness according the existing clinical guidelines [11]. In an attempt to verify if PQM can be used to predict accuracy of septal wall thickness measurement algorithm, we visualized the error in thickness measure and PQM as a scatter plot in Fig. 4(c). We observed that on images with low PQM, the probability of error in septum thickness measurement will be high and *vice versa*.
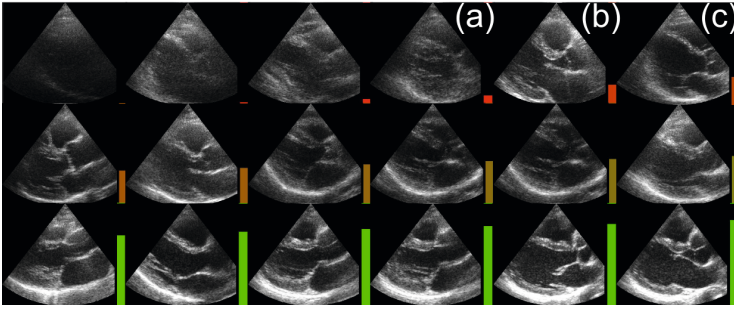
Fig. 5. PQM, illustrated as vertical bars, computed on few test images. The height of the bar is proportional to PQM. Note that PQM is low for (a), (b), and (c), which represent the bad quality PLAX images illustrated in Fig. 1. As the anatomical components of the heart become progressively more visible, PQM increases.

Thus, one could deploy automatic segmentation algorithms only when PQM is greater than a preset threshold, thereby, providing an upper bound on the segmentation error.

Visual Assessment of PQM: Fig. 5 visualizes PQM as vertical bars. The height of the bar is proportional to the PQM computed for an image. Typical poor quality PLAX echocardiograms are shown in sub-figures (a), (b) and (c). The contrast between the blood pool and the structures of interest is low in (a). Pericardium is not visualized in (b). Sub-figure (c) illustrates the case where septum and pericardium are not parallel to each other, and this signifies an incorrect scan plane. Note that PQM is low for all three cases. For images with better quality, one sees an increase in PQM.

4   Conclusions and Discussion

In this paper, we proposed a solution to the previously unstudied problem of computing a quality metric for Parasternal Long AXis (PLAX) view echocardiograms. Our algorithm checks for the presence of expected anatomical structures in a PLAX image (Septum, mitral valve and pericardium) using Generalized Hough Transform (GHT). Based on the evidence GHT accumulates during the search, we output PQM that seems to correlate well with an expert's rating and CNR metric. We also observe that for images with high PQM, error in septum thickness measurement algorithm decreases.

Because the expected structures in a PLAX image are relatively large, we observed that we could subsample the image to 50×50 pixels, and still see the structures of interest. At this resolution, we could obtain a processing speed of 35 frames a second in a Intel i7 2.67 GHz processor. The implementation was done in C++ using the Insight Segmentation and Registration Toolkit [12]. The results illustrated in this paper have been performed by resizing images to 50×50 pixels.

The proposed quality metric is the first step in guiding the user into acquiring the right scan plane. The user will receive real-time feedback to the acquired image, thereby, motivating the user to acquire a better quality image. Although, the proposed method is specific for PLAX images, we perceive easy extension to other anatomies using an appropriate atlas and a suitable contrast enhancement technique.

References

1. Alonso-Fernandez, F., Fierrez-Aguilar, J., Ortega-Garcia, J.: A review of schemes for fingerprint image quality computation. In: Cost-275 Workshop, pp. 3–6 (2005) 1
2. Goodsitt, M., Carson, P., Witt, S., Hykes, D., Kofler, J.: Real-time B-mode ultrasound quality control test procedures. Medical Physics 25, 1385–1406 (1998) 1
3. Duda, R., Hart, P.: Use of the Hough transformation to detect lines and curves in pictures. Communication ACM 15, 11–15 (1972) 2, 3
4. Ballard, D.: Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition 13, 111–122 (1981) 2, 3
5. Boukerroui, D., Baskurt, A., Noble, J., Basset, O.: Segmentation of ultrasound images. Pattern Recognition Letters 24, 779–790 (2003) 3
6. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998) 3
7. Antiga, L.: Generalizing vesselness with respect to dimensionality and shape. The Insight Journal (2007) 3
8. Gonzalez, R., Woods, R.: Digital image processing, ch. 9 (1993) 4
9. Rohlfing, T., Maurer, C.: Shape-based averaging. IEEE Transactions on Image Processing 16, 153–161 (2007) 4
10. Subramanian, N., Padfield, D., Thiruvenkadam, S., Narasimhamurthy, A., Frigstad, S.: Automated Interventricular Septum Thickness Measurement from B-Mode Echocardiograms. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 510–517. Springer, Heidelberg (2010) 7
11. Lang, R., Bierig, M., Devereux, R., Flachskampf, F., Foster, E., Pellikka, P., Picard, M., Roman, M., Seward, J., Shanewise, J., Solomon, S., Spencer, K., Sutton, M., Stewart, W.: Recommendations for chamber quantification. European Journal of Echocardiography 7, 79–108 (2006) 7
12. Ibáñez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. Kitware (2003) 8

Hemodynamic Assessment of Pre- and Post-operative Aortic Coarctation from MRI

Kristóf Ralovich[1,2], Lucian Itu[2,3], Viorel Mihalef[2], Puneet Sharma[2],
Razvan Ionasec[2], Dime Vitanovski[2,4], Waldemar Krawtschuk[2,4],
Allen Everett[5], Richard Ringel[5], Nassir Navab[1], and Dorin Comaniciu[2]

[1] Technical University of Munich, Germany
[2] Siemens Corporation, Corporate Research and Technology, USA
[3] Transilvania University of Brasov, Romania
[4] Friedrich-Alexander University Erlangen-Nuremberg, Germany
[5] The Johns Hopkins Hospital, USA

Abstract. Coarctation of the aorta (CoA), is a congenital defect characterized by a severe narrowing of the aorta, usually distal to the aortic arch. The treatment options include surgical repair, stent implantation, and balloon angioplasty. In order to evaluate the physiological significance of the pre-operative coarctation and to assess the post-operative results, the hemodynamic analysis is usually performed by measuring the pressure gradient ($\triangle P$) across the coarctation site via invasive cardiac catheterization. The measure of success is reduction of the ($\triangle P > 20mmHg$) systolic blood pressure gradient. In this paper, we propose a non-invasive method based on Computational Fluid Dynamics and MR imaging to estimate the pre- and post-operative hemodynamics for both native and recurrent coarctation patients. High correlation of our results and catheter measurements is shown on corresponding pre- and post-operative examination of 5 CoA patients.

1   Introduction

Coarctation of the aorta (CoA) accounts for $5 - 8\%$ of the $8/1,000$ congenital heart disease (that is $4-6/10,000$) live births [1] in the USA. CoA is a congenital defect characterized by a severe narrowing of the aorta, usually distal to the aortic arch. Patients born with CoA require lifelong medical/surgical care, that includes invasive and non-invasive imaging, drug therapy and if the CoA recurs, invasive catheterization or surgical intervention to reduce the blood pressure in the ascending aorta.

Pre-operative evaluation of CoA severity relies predominately on non-invasive arm/leg blood pressure gradients or if anatomy makes that comparison not feasible, estimation by Doppler ultrasonography. The clinical gold-standard is obtained by invasive cardiac catheterization to measure ($\triangle P$) across the coarctation site. Recently, Doppler and phase contrast (PC) MRI based methods have been proposed for a non-invasive estimation of $\triangle P$ [2] by using simplified

relationships (e.g. modified Bernoulli equation) between flow and pressure. However, such approaches often overestimate the pressure difference and do not provide other hemodynamic parameters (such as wall-shear stress, vorticity, etc.) that characterize the complex flow-field associated with CoA, which could potentially be used for diagnosis and treatment planning.
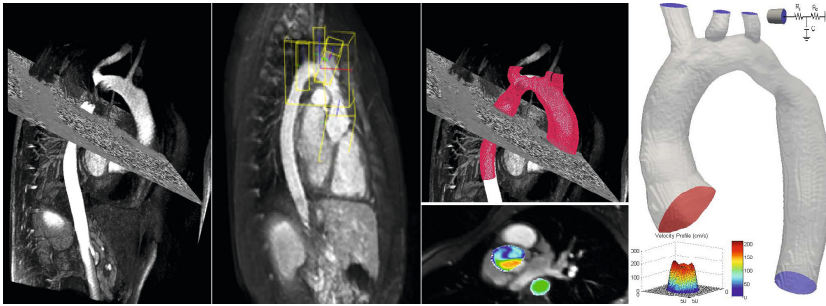
Fig. 1. Left: Overlay of angiogram and PC-MRI **Left Middle**: Bounding boxes for parts of the vessel tree, retrieved by learning based detection **Right Middle**: Segmented vessel tree and delineated aortic flow profiles **Right**: Overview of boundary conditions applied, measured flow profile at AAo, Windkessel model at carotid and descending outlets

Multiple groups have investigated CoA hemodynamics through simulation. Coupling the aorta with a lumped parameter model of the left ventricle, [3] applies realistic inflow boundary conditions, but only studies two patients. Outlet conditions are treated in [4], but employing simple heuristics at the supra aortic branches, also pre-operative simulation is not compared with post repair of the same patient. [5] applies detailed boundary conditions and accurate simulation procedure, but imaging data was specifically acquired for the study, and potentially difficult to reproduce in standard exams.

To address these issues, we propose an MR image-based pre-processing and hemodynamics simulation pipeline for thoracic aortic investigation. A patient-specific model of the aorta and supra-aortic arteries is automatically estimated using a discriminative learning-based method. Based on the lumen geometry, we introduce a CFD setup employing personalized boundary conditions. The simulation provides dense 3D+time velocity and pressure maps. We applied the method to existing data of CoA patients of multiple hospitals enrolled in a FDA authorized clinical trial [1]. Blood pressure computation is validated against invasive catheterization on 5 cases. Strengths of our method are 1) end-to-end computational pipeline that may be integrated within the clinical workflow, 2) the ability to work on clinical images acquired within existing protocols and 3) reproducible simulation results of good initial agreement.

2   Proposed Method

2.1   Estimation of the Patient-Specific Lumen

An accurate geometrical representation of the lumen boundaries is essential for subsequent simulations. To facilitate reproducibility of our computation, we developed a robust learning-based method to estimate a model of the aorta and supra-aortic arteries from contrast enhanced MR angiograms (CE-MRA). The thoracic aorta and main branches are represented as multiple parts: aortic root, aortic arch, walls of ascending and descending aorta, and the supra-aortic arteries (SAoA) - brachiocephalic trunk, left common carotid artery, left subclavian artery (illustrated on Fig. 1, **Left Middle**).

In a first step, the pose of each part is estimated following a hierarchical scheme that allows for the utilization of anatomical constraints. In image I, each pose θ is parametrized as a 3D affine transformation. We formulate the estimation as a multi-object detection problem and learn the posterior probability $p(\theta|I)$ using the Probabilistic Boosting Tree [6]. As the aortic arch pose estimation produced most accurate results, we exploit its anatomical proximity to the SAoA, and constrain the pose estimation of the former by learning the variation in their relative distances from the available training set. Constrained by the estimated poses, each part is initialized with a corresponding mean model constructed by employing statistical shape analysis [7]. A lumen detector is trained using the PBT and Haar-like features to drive the deformation of the initial model towards the actual boundary in the image. The final model is obtained by merging the separately estimated parts using a sequence of forward and backward projection to/from Eulerian representation to retrieve the composited Lagrangian arterial tree geometry.

2.2   Estimation of the Patient-Specific Blood Flow from PC-MRI

To quantify the subject's measured blood flow conditions, patient-specific flow profiles over the entire cardiac cycle are extracted at the aortic in- and outflow from the velocity encoded 2D PC-MRI Cine images. These sequences contain through-plane blood flow measurements in an oblique arrangement, intersecting the aorta twice: at the ascending aorta (AAo) above the valve and in the region of the descending aorta (DAo). Given a centerline of the aorta calculated from the segmentation, delineation of the lumen boundary on the MR image is initialized using graph cuts, and physiological radii constraints. The single time point segmentation is then tracked throughout the cardiac cycle similar to [8]. Inside each patch, sampling of the PC-MR image is performed at the pixel centers to obtain velocity values over the entire cardiac cycle. This velocity field is integrated over the area of the segmented contour to estimate the aortic blood flow rate.

2.3   Non-invasive Estimation of Personalized Boundary Conditions

In order to perform a patient-specific simulation, the inlet and outlet boundary conditions are personalized based on the non-invasive measurements. Given the high computational cost associated with unsteady 3D flow simulations, we use an axisymmetric reduced-order model of aortic hemodynamics for a quick estimation of patient-specific parameters that are subsequently used for a personalized 3D flow simulation.

The inlet boundary condition is specified by imposing the time-varying flow rate in the ascending aorta obtained from the PC-MRI data. For the outlet boundary condition, we use a 3-element Windkessel model to specify the downstream resistance and compliance of the vessels that are not explicitly modeled in the flow computations. At each of the four outlets (SAoA and DAo), the windkessel model consists of two resistances (R_p and R_d) and one compliance (C). R_p is proximal resistance, R_d the distal resistance and C the total compliance of the downstream circulation for each outlet. At each outflow, the total resistance R_t, (where $R_t = R_p + R_d$) is estimated from the mean arterial pressure (P_{MA}) and average flow (Q) by using the relation: $R_t = \frac{P_{MA}}{Q}$. For the non-invasive estimation of P_{MA} in the AAo, we use the systolic and diastolic cuff pressures, together with the heart rate: $P_{MA} = P_{DB} + \left[\frac{1}{3} + (Hr \cdot 0.0012)\right](P_{SB} - P_{DB})$, where P_{DB} and P_{SB} are the diastolic and systolic blood pressures and Hr is the heart rate.

The average flows at the ascending (Q_{AAo}) and the descending aorta (Q_{DAo}) are obtained from the PC-MRI data. Hence, the total flow in the three remaining outflow vessels (Q_{up}) can be obtained by: $Q_{up} = Q_{AAo} - Q_{DAo}$. As has been shown previously [9], the total flow in the first few branches, starting from the aortic root, is distributed proportionally to the square of the vessel radius, i.e. $Q_i = \frac{Q_{up} \cdot r_i^2}{\sum_{i=1}^{3} r_i^2}$ where r_i is the vessel radius at the outflow of upper branch i. Since the pressure difference between the ascending aorta and the three upper branches is insignificant, we use the same average pressure to estimate the total resistance at each upper outlet branch: $(R_t)_i = \frac{P_{MA}}{Q_i}$. For the descending aorta, the assumption that P_{MA} is the same as for the ascending aorta does not hold true, since the coarctation induces a significant pressure drop. The coarctation introduces a flow-dependent resistance ($R_s(Q)$) and thus, the total resistance, which represents the sum of the resistance of the coarctation and of the outflow model, is estimated as follows: $(R_t)_{DAo} + R_s(Q) = \frac{P_{MA}}{Q_{DAo}}$. The resistance of the coarctation is estimated with a semi-empirical model: $R_s(Q) = \frac{\mu \cdot K_v}{2 \cdot \pi \cdot r_0^3} +$ $\frac{\rho \cdot K_t}{2 \cdot A_0^2}\left(\frac{A_0}{A_s} - 1\right)^2 Q_{DAo}$ where r_0 is the proximal radius of the coarctation, A_0 and A_s are the proximal and minimum cross-sectional areas, μ is the dynamic viscosity, ρ is the blood density and K_v and K_t are two constants which represent the viscous and turbulent losses of energy. Q_{DAo}, the measured average flow rate through the descending aorta. Using the last two equations the total resistance at the descending aorta is then computed. After the computation of the total resistances, the last step is to compute R_p and R_d. R_p at each outflow is equal to the characteristic resistance of the vessel (in order to minimize the reflections),

which is computed by the expression $(R_p)_i = \frac{1}{\Pi \cdot r_i^2} \sqrt{\frac{2 \cdot \rho \cdot E \cdot h}{3 \cdot r_i}}$, as described in [10]. R_d is then computed by $R_d = R_t - R_p$. For estimating the total compliance we use the method in [11] and redistribute the individual compliances at each outlet. Once the windkessel parameters are estimated, no further tuning is performed to run the simulations.

2.4   Patient-Specific 3D CFD Simulations

To obtain comprehensive flow information in three dimensions we solve the full 3D Navier-Stokes equations in the luminal aortic domain calculated, using the personalized outflow boundary conditions for pressure. We use an embedded boundary method for automatic transfer of the tagged STL triangular lumen mesh into a cartesian domain. The embedding function is a signed-distance function computed using the Closest Point Transform [12]. The computational domain cells are tagged based on their relation with the inlet triangular mesh as follows: *Exterior* (no computation is taking place), *Interior* (computation is taking place), *Inlet*, *Outlet* and *Wall* (appropriate boundary conditions are imposed). The *Inlet* and *Wall* cells are all interior to the domain, while the *Outlet* cells are situated on the domain boundaries, by extending the lumen of each vessel in its centerline direction [13] until it reaches the cartesian boundary. Our embedded boundary Navier-Stokes solver uses a fractional step method [14] that computes in a first step an intermediate velocity field, using the nonlinear advection-diffusion equation for velocity, and then projects the intermediate velocity onto the field of divergence free and tangent to the vessel boundary vector fields. For the velocity advection we use second-order upwind, Van-Leer slope limiting methods, while for the diffusion force components we use a semi-implicit approach as in [15] which is first order accurate and unconditionally stable in 3D. We solve the pressure projection Poisson equation using an efficient implicit multi-grid preconditioned conjugate gradient solver. The boundary conditions for the velocity are Dirichlet in the *Inlet* cells, no-slip (Dirichlet) in the *Wall* cells, and Neumann in the *Outlet* cells [16]. We use a variable-in-time flat inlet velocity profile, and outlet pressure boundary conditions provided by the axisymmetric 1D simulations. The blood density and viscosity are set to literature based values for healthy individuals ($1.05g/cm^3$ and $4mPa \cdot s$).

3   Experiments and Validation

3.1   Segmentation

The accuracy of our method is demonstrated on a set of 212 3D images with wide range of morphological and pathological variations. Data was retrospectively collected from multiple hospitals participating in the COAST Trial [1], acquired using heterogeneous MR protocols/sequences with scanners from all three major vendors. Each volume in the data set is associated with an annotation (from manual operators) which is considered as ground truth for training and testing.

The fused aortic vessel tree segmentation accuracy was evaluated by using the point-to-mesh metric (Table 1a). Average detection time for all parts is in the range of $2-3$ minutes on an Intel Core i7 laptop.

From a clinical perspective, the quantitative capabilities of our system are demonstrated on 32 patients with aortic anomalies (age: 5-36 years, 17 with CoA and 15 with bicuspid aortic valve and dilated AAo) by comparing a set of morphological measurements[17] automatically derived from our segmentation to measurements manually performed by our cardiologist collaborators. The aortic min and max diameter were measured at five location: aortic sinus (**AS**), sino-tubular junction (**STJ**), ascending aorta (**AAO**), transverse arch (**TA**), and descending aorta (**DA**). Statistical results correlated ($p < 0.001$, $r = 0.94$) between model-based and manually measured min and max diameters. Table 1b summarizes the mean measurement errors for each segment separately.

3.2   Parameter Estimation, Simulation Results and Validation

In order to demonstrate the proposed method for non-invasive hemodynamic assessment, we investigated 5 patient datasets with native and/or recurrent coarctation that involved the aortic isthmus. The stenoses received repair through balloon angioplasty and stenting.

Using the above methods the patient-specific geometric vessel wall model and corresponding time-resolved flow profiles were estimated from MR images. To make sure that segmentation errors (e.g. results of image quality due to MR signal drop-out inside the metallic stent for post-op) do not influence the simulation outcomes, the vessel tree geometry was reviewed by a manual operator in all experiments.

Given the patient-specific anatomy, measured flow rates at the AAo and DAo, and systolic and diastolic blood pressures and heart rate, we performed a non-invasive parameter estimation of the boundary conditions for each patient. The results obtained for the non-invasive pressure gradients are summarized in Table 1c, together with the invasive pressures obtained from cardiac catheterization. The pressure differences between the AAo-DAo and TAA-DAo are determined at the time-instant when the flow rate through the descending aorta is maximal.

Table 1. a) Segmentation accuracy (mm) and **b**) comparison between manual and model-based clinical measurements (mm). **c**) Comparison of the pressure obtained from invasive catheterization and our proposed non-invasive method: systolic blood pressure gradients ($mmHg$) between AAo-DAo and transverse aortic arch (TAA)-DAo.

(a)

	Mean (Std)	Median
aorta	2.29 ± 1.74	1.95
brachiocephalic	3.40 ± 1.89	2.90
left common	4.59 ± 3.58	3.16
left subclavian	4.64 ± 3.33	3.06
supra-aortic	4.21 ± 2.90	3.04
complete model	3.24 ± 2.32	2.49

(b)

	AS	STJ	AAO
min	1.61 ± 0.9	2.07 ± 1.5	1.61 ± 1.9
max	1.56 ± 1.3	1.28 ± 1.0	1.56 ± 1.3
	TA	DA	
min	1.70 ± 1.2	0.8 ± 0.5	
max	1.34 ± 1.1	0.92 ± 0.6	

(c)

Patient	$\triangle P$ AAo-DAo	$\triangle P$ TAA-DAo
#1	55/58.5	53/56
#3	8/6.9	8/7.7
#4	30/17.3	28/24.3
#5	14/15	18/13.2
#6	39/6.7	43/6.1

4   Discussion

As can be seen from the results, the proposed method performs well for most of the cases. For patients #1, #3 and #5 the simulation reproduces the catheterization within an acceptable margin. During the review of our results for case #6 we observed that an incorrect PC-MR acquisition plane (intersecting the aortic valve and left ventricular outflow tract instead of the AAo) that results in an erroneous inflow initialization, thus we just included the results for consistency. The discrepancy in #4 is due to aortic compliance as the DAo peak flow phase is delayed compared to AAo peak blood pressure.

The results from the 3D simulations for two patients are presented in Figure 2. There is a significant pressure gradient across the coarctation at peak systole, which gradually disappears towards the end of diastole. A volumetric visualization of the velocity magnitude at peak systole, late systole and end diastole are shown in the three figures at the right. The high velocity jet in the stenosis region is clearly visible as expected. Similar methodology can also be applied to the post-operative data, by taking into consideration the modified wall-stiffness introduced from the stent implantation. This is part of the ongoing work, and the preliminary results are shown in the bottom row of Figure 2. Here, the pressure gradients between AAo-DAo and TAA-DAo have been partially restored to normal values. A similar effect can also be noticed in the flow patterns: highest velocity in the aortic arch, and reduced Reynolds number.
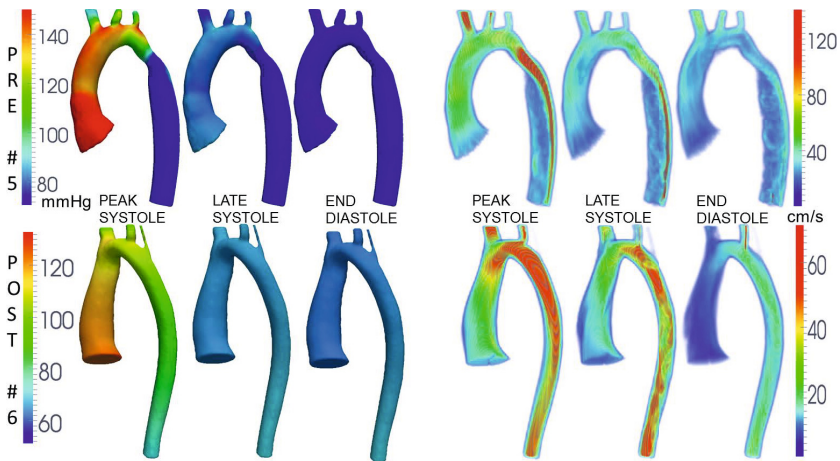
Fig. 2. Left:Blood pressure mapped on lumen boundary, **Right**:Volume rendered velocity magnitude (for cases pre-operative #5, post-operative #6)

References

1. Ringel, R.E., Jenkins, K.: Coarctation Of the Aorta Stent Trial (COAST) (2007)
2. Coogan, J.S., et al.: Computational fluid dynamic simulations of aortic coarctation comparing the effects of surgical- and stent-based treatments on aortic compliance and ventricular workload. Catheterization and Cardiovascular Interventions (2011)
3. Kim, H., Vignon-Clementel, I., Figueroa, C., LaDisa, J., Jansen, K.E.A.: On coupling a lumped parameter heart model and a three-dimensional finite element aorta model. Annals of Biomedical Engineering (2009)
4. Valverde, I., et al.: Predicting hemodynamics in native and residual coarctation: preliminary results of a rigid-wall computational-fluid-dynamics model (rw-cfd) validated against clinically invasive pressure measures at rest and during pharmacological stress. Journal of Cardiovascular Magnetic Resonance (2011)
5. LaDisa, J.F.J., Alberto Figueroa, C., Vignon-Clementel, I.E., Kim, H.J., Xiao, N., Ellwein, L.M., Chan, F.P., Feinstein, J.A., Taylor, C.A.: Computational simulations for aortic coarctation: representative results from a sampling of patients. Journal of Biomechanical Engineering (2011)
6. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: Proc. ICCV, vol. 2 (2005)
7. Yefeng, Z., et al.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. IEEE Transactions on Medical Imaging (2008)
8. Jolly, M.P., Guetter, C., Guehring, J.: Cardiac segmentation in mr cine data using inverse consistent deformable registration. In: ISBI (2010)
9. Zamir, M., Sinclair, P., Wonnacott, T.H.: Relation between diameter and flow in major branches of the arch of the aorta. Journal of Biomechanics (1992)
10. Olufsen, M.S., Peskin, C.S., Kim, W.Y., Pedersen, E.M., Nadim, A., Larsen, J.: Numerical simulation and experimental validation of blood flow in arteries with structured-tree outflow conditions. Annals of Biomedical Engineering 28 (2000)
11. Stergiopulos, N., Westerhof, B.E., Westerhof, N.: Total arterial inertance as the fourth element of the windkessel model. American Journal of Physiology (1999)
12. Mauch, S.: Efficient algorithms for solving static Hamilton-Jacobi equations. PhD Thesis (2003)
13. Ralovich, K., Ionasec, R., Mihalef, V., Sharma, P., Georgescu, B., Everett, A., et al.: Computational fluid dynamics framework for large-scale simulation in pediatric cardiology. In: Computational Biomechanics for Medicine VI (2011)
14. Bell, J.B., Colella, P., Glaz, H.M.: A second-order projection method for the incompressible Navier-Stokes equations. J. Comput. Phys. (1989)
15. Li, J., Renardy, Y., Renardy, M.: Numerical simulation of breakup of a viscous drop in simple shear flow through a volume-of-fluid method. Phys. Fluids (2000)
16. Mihalef, V., Ionasec, R.I., Sharma, P., Georgescu, B., Voigt, I., et. al.: Patient-specific modelling of whole heart anatomy, dynamics and haemodynamics from four-dimensional cardiac ct images. In: Interface Focus. (2011)
17. Pasqua, A., et al.: Comparison of contrast and noncontrast magnetic resonance angiography for quantitative analysis of thoracic arteries in young patients with congenital heart defects. Ann. Pediatr. Cardiology 4(1) (2011)

Linear Invariant Tensor Interpolation Applied to Cardiac Diffusion Tensor MRI

Jin Kyu Gahm[1,2], Nicholas Wisniewski[3], Gordon Kindlmann[5],
Geoffrey L. Kung[1], William S. Klug[4], Alan Garfinkel[3], and Daniel B. Ennis[1]

[1] Department of Radiological Sciences, UCLA, CA 90095, USA
[2] Department of Computer Science, UCLA, CA 90095, USA
[3] Department of Medicine (Cardiology), UCLA, CA 90095, USA
[4] Department of Mechanical and Aerospace Engineering, UCLA, CA 90095, USA
[5] Computer Science Department, University of Chicago, Chicago, IL 60637, USA
gahmj@ucla.edu

Abstract. *Purpose:* Various methods exist for interpolating diffusion tensor fields, but none of them linearly interpolate tensor shape attributes. Linear interpolation is expected not to introduce spurious changes in tensor shape. *Methods:* Herein we define a new linear invariant (LI) tensor interpolation method that linearly interpolates components of tensor shape (tensor invariants) and recapitulates the interpolated tensor from the linearly interpolated tensor invariants and the eigenvectors of a linearly interpolated tensor. The LI tensor interpolation method is compared to the Euclidean (EU), affine-invariant Riemannian (AI), log-Euclidean (LE) and geodesic-loxodrome (GL) interpolation methods using both a synthetic tensor field and three experimentally measured cardiac DT-MRI datasets. *Results:* EU, AI, and LE introduce significant microstructural bias, which can be avoided through the use of GL or LI. *Conclusion:* GL introduces the least microstructural bias, but LI tensor interpolation performs very similarly and at substantially reduced computational cost.

1 Introduction

Diffusion tensor magnetic resonance imaging (DT-MRI) [1] is a technique that permits the non-destructive evaluation of the self-diffusion tensor ($\mathbf{D}$) of water within small volumes of soft tissues. The measured diffusion tensor can be used to characterize local microstructural tissue properties, including diffusive shape and microstructural orientation. Diffusion tensor shape and orientation properties are important components of computational models of cardiac mechanics and electrophysiology that require closely spaced nodes that do not necessarily lie at lattice points. DT-MRI data are, however, acquired at lattice points within a three-dimensional imaging volume, therefore tensor interpolation methods are needed.

Each diffusion tensor is a three-dimensional rank-2 symmetric, positive definite tensor that can be decomposed into a system of eigenvalues (λ_i) and eigenvectors ($\boldsymbol{e}_i$), which correspond to tensor shape and orientation respectively.

In particular, tensor shape can be described by families of tensor invariants [2, 3], which saliently decompose tensor shape into the magnitude-of-isotropy (tensor trace, determinant or norm), magnitude-of-anisotropy (fractional or relative anisotropy) and mode-of-anisotropy (transversely isotropic vs. orthotropic) components. The primary eigenvector corresponds to the direction of fastest diffusion, which has been shown to align with the long axis of the myocytes that comprise the heart [4]. The secondary and tertiary eigenvectors correspond to the crossfiber-within-sheet direction and the sheet normal direction, respectively [4]. Because these shape and orientation characteristics directly correspond to microstructural features that are observed with histology, it is judicious to linearly interpolate all of these tensor properties directly.

The simplest tensor interpolation method is the Euclidean (EU) method, but it suffers from the swelling effects due to non-monotonic interpolation of tensor determinant (DET), and does not preserve the positive definiteness of diffusivity. The Riemannian approaches [5–8] overcome this problem, and more recently the log-Euclidean (LE) method [9] has been shown to be a computationally efficient approximation to the affine-invariant (AI) Riemannian approach [8]. Kindlmann *et al.* [10] proposed a geodesic-loxodrome (GL) approach that guarantees monotonic interpolation of orthogonal tensor invariants, and demonstrated that the EU, AI and LE approaches fail to monotonically interpolate all the tensor invariants including tensor trace (TR), fractional anisotropy (FA) and tensor mode (MODE). The geodesic-loxodrome approach, however, is computationally expensive, and monotonic interpolation of the tensor invariants needs to be evaluated using experimentally measured DT-MRI data.

Recent studies have examined different methods to interpolate separately tensor shape and orientation [11, 12]. Bi *et al.* [11] proposed a method to linearly interpolate eigenvalues and rotation angles between tensor orientations. This results in monotonic interpolation only for TR. Bi *et al.* did not provide a way to resolve the sign ambiguity of the eigenvectors. Yang *et al.* [12] proposed a method to resolve the sign ambiguity problem by finding the minimum rotation path between tensor orientations, but the minimum rotation path may not be the best way to resolve the sign ambiguity problem.

Firstly, we propose a new linear invariant (LI) tensor interpolation method, which linearly interpolates components of tensor shape (tensor invariants). We also define for the first time the necessary mathematics for converting the tensor invariants to eigenvalues, which enables recapitulation of the interpolated tensor from the linearly interpolated tensor invariants and the eigenvectors of a linearly interpolated tensor. The LI tensor interpolation method is simple to implement, fast, and perfectly commutative. Secondly, we determine which tensor interpolation scheme introduces the least microstructural bias to the shape and orientation of the interpolated tensors. To do so the LI tensor interpolation method is compared to the EU, AI, LE and GL methods of tensor interpolation using both a synthetic tensor field that reflects important myocardial tensor field attributes, and three experimentally measured DT-MRI datasets from rabbit, pig and human hearts.

2   Theory

A tensor invariant set composed of TR (K_1, magnitude-of-isotropy), FA (R_2, magnitude-of-anisotropy) and MODE (R_3, mode-of-anisotropy) fully decompose the shape of a tensor $\mathbf{D}$ defined by [2, 3]:

$$K_1(\mathbf{D}) = \mathrm{tr}(\mathbf{D}), \quad R_2(\mathbf{D}) = \sqrt{\tfrac{3}{2}}|\mathbf{D}|/|\widetilde{\mathbf{D}}|, \quad R_3(\mathbf{D}) = 3\sqrt{6}\det(\widetilde{\mathbf{D}}/|\widetilde{\mathbf{D}}|) , \qquad (1)$$

where tr() and det() are the trace and determinant operators respectively, $|\mathbf{D}|$ represents the magnitude (Frobenius norm) of $\mathbf{D}$ defined by $|\mathbf{D}| = \sqrt{\mathrm{tr}(\mathbf{D}\mathbf{D}^T)}$, and $\widetilde{\mathbf{D}}$ represents the anisotropic (deviatoric) part of $\mathbf{D}$ defined by $\widetilde{\mathbf{D}} = \mathbf{D} - \mathrm{tr}(\mathbf{D})\mathbf{I}/3$. Linear invariant (LI) interpolation of tensor $\mathbf{C}_{\mathrm{LI}}$ from tensors $\mathbf{A}$ and $\mathbf{B}$ with weighting coefficient $t \in [0, 1]$ starts by linearly interpolating the tensor invariants:

$$K_1(\mathbf{C}_{\mathrm{LI}}) = (1 - t)K_1(\mathbf{A}) + tK_1(\mathbf{B}), \quad R_j(\mathbf{C}_{\mathrm{LI}}) = (1 - t)R_j(\mathbf{A}) + tR_j(\mathbf{B}) , \quad (2)$$

for $j = 2, 3$. Without derivation we define the mathematics for converting the tensor invariants into the eigenvalues by solving the cubic characteristic polynomial for a tensor:

$$\lambda_i = \tfrac{1}{3}K_1 + \frac{2K_1 R_2}{3\sqrt{3 - 2R_2^2}} \cos\left(\frac{\arccos(R_3) + P_i}{3}\right) , \qquad (3)$$

where $P_i = 0, -2\pi, 2\pi$ for $i = 1, 2, 3$. Then Eq. 3 permits converting the tensor invariants of $\mathbf{C}_{\mathrm{LI}}$ into the eigenvalues $\lambda_i(\mathbf{C}_{\mathrm{LI}})$.

To define the eigenvectors for $\mathbf{C}_{\mathrm{LI}}$, we use linear (EU) tensor interpolation $\mathbf{C}_{\mathrm{EU}} = (1 - t)\mathbf{A} + t\mathbf{B}$, then decompose $\mathbf{C}_{\mathrm{EU}}$ into the eigenvector and eigenvalue matrices $\mathbf{R}_{\mathrm{EU}}$ and $\mathbf{\Lambda}_{\mathrm{EU}}$ where $\mathbf{C}_{\mathrm{EU}} = \mathbf{R}_{\mathrm{EU}}\mathbf{\Lambda}_{\mathrm{EU}}\mathbf{R}_{\mathrm{EU}}^T$. We can use AI, LE or GL tensor interpolation, but EU is the simplest and fastest, and introduces a similar bias in tensor orientation recovery, as will be shown later in Section 4.

Finally the interpolated tensor $\mathbf{C}_{\mathrm{LI}}$ is constructed using the eigenvalue matrix $\mathbf{\Lambda}_{\mathrm{LI}} = \mathrm{diag}(\lambda_i(\mathbf{C}_{\mathrm{LI}}))$ from the linearly interpolated *tensor invariants*, and the eigenvector matrix $\mathbf{R}_{\mathrm{EU}}$ from the linearly interpolated *tensor*:

$$\mathbf{C}_{\mathrm{LI}} = \mathbf{R}_{\mathrm{EU}}\mathbf{\Lambda}_{\mathrm{LI}}\mathbf{R}_{\mathrm{EU}}^T . \qquad (4)$$

3   Methods

Synthetic Tensor Field. Using the EU, LE, GL and LI tensor interpolation methods, bilinear interpolation was performed between tensors that approximate the sheet shape of $(K_1, R_2, R_3) = (7, 0.6, 0.5)$ and fiber shape of $(6, 0.7, 1)$ observed in cardiac DT-MRI data, and range of tensor orientations.

Real Cardiac DT-MRI Acquisition. The rabbit heart DT-MRI data was acquired in a formalin fixed rabbit heart using a 7T Bruker Biospin scanner and a 3D fast spin echo sequence. Five non-diffusion weighted and twenty-five diffusion

weighted (b-value $= 1000\,\mathrm{s/mm}^2$) imaging volumes were used to estimate the local **D** without zero padding and with linear regression. The imaging parameters were TE/TR $= 29.1/550\,\mathrm{ms}$, RARE factor two, FOV $= 35 \times 35 \times 35\,\mathrm{mm}$, and a $96 \times 96 \times 72$ encoding matrix – resulting in $365 \times 365 \times 528\,\mu m$ spatial resolution. The porcine heart DT-MRI data was acquired using 2D multislice readout segmented EPI, similar encoding directions and reconstruction method with the following imaging parameters: TE/TR $= 80/6800\,\mathrm{ms}$, FOV $= 150 \times 150 \times 129\,\mathrm{mm}$, and an $150 \times 150 \times 43$ encoding matrix – resulting in $1 \times 1 \times 3\,\mathrm{mm}$ spatial resolution. The high-resolution human heart DT-MRI data was downloaded from Johns Hopkins University [13]. The FOV was $110 \times 110 \times 110\,\mathrm{mm}$, the encoding matrix size was $256 \times 256 \times 134$, and the spatial resolution was $0.430 \times 0.430 \times 1.0\,\mathrm{mm}$.

Interpolation Evaluation. Each DT-MRI volume was segmented to identify the myocardium using thresholding and morphologic operations on the non-diffusion weighted image volume. To evaluate each interpolation method the measured ("truth") tensor volume was first downsampled in each direction by a factor of 2 for the rabbit and porcine heart data, and by a factor of 4 for the high-resolution human heart data. Subsequently tensors were trilinearly interpolated at the positions of the removed tensors using the remaining data. This permits a direct, paired comparison of the interpolated tensors to the "truth" tensors using data de-correlation and bootstrap statistics (see below). This comparison was made for six tensor scalar measures (TR, FA, MODE, DET, angle difference between primary eigenvectors, and log-Eugclidean tensor distance [9]) computed at each location of the interpolated tensors using each of the tensor interpolation methods.

Tensor Statistics. The distributions of the six tensor scalar measures contain correlated data, are non-Gaussian, and have non-uniform variances. The use of ANOVA and t-test statistics, however, requires that the data in each population are not correlated, are Gaussian distributed (negligible skewness, kurtosis, etc.), and have similar variances; hence de-correlation and bootstrap methods are required.

De-correlation. The population of each tensor scalar measure was spatially decorrelated by computing the autocorrelation (AC) length for every dimension using the fully sampled data and the mask. For each of the $x-$, $y-$ and $z-$directions, all lines having at least four continuous myocardial points were found within the mask. For each line, the data values of the line were subtracted from their average, and then the AC sequence was computed. The AC length, which is the lag value at the first zero-crossing of the AC curve, was computed. The interpolated and original tensor data were conservatively decimated by the minimum integer value greater than or equal to all the median AC lengths for the tensor scalar measures in every dimension in order to spatially de-correlate the data.

Bootstrap Statistics. A paired comparison of each scalar tensor measure between the de-correlated interpolated tensors and the de-correlated original "truth" tensors was made using bootstrap methods. The population of paired differences between the scalar tensor measures (interpolated minus "truth" values)

(a) EU interpolation

(b) LE interpolation

(c) GL interpolation

(d) LI interpolation

(e) TR

(f) FA

(g) MODE

(h) DET
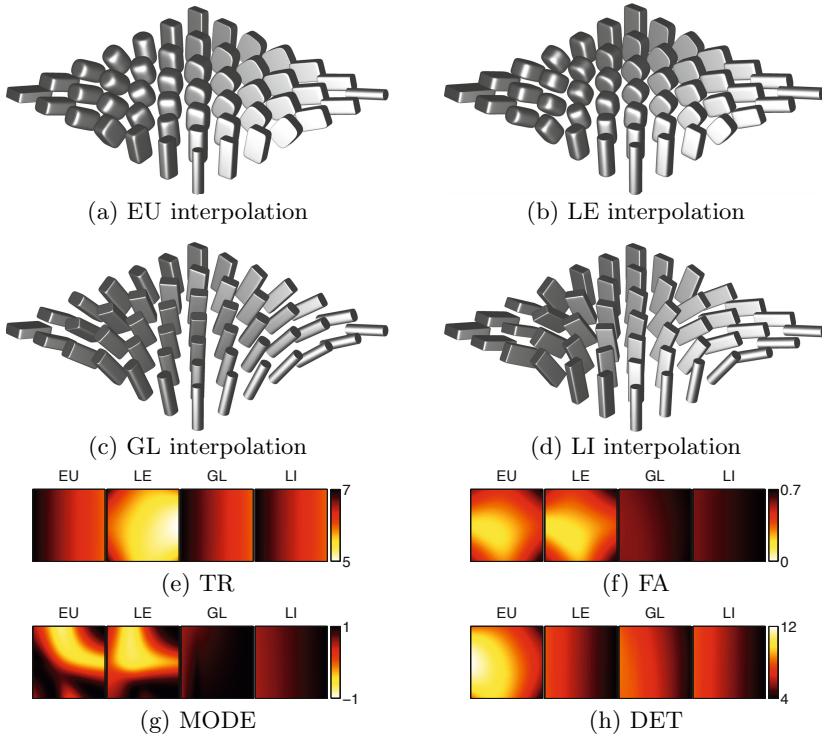
Fig. 1. Superquadric glyph rendering of the tensor field obtained from Euclidean (a), log-Euclidean (b), geodesic-loxodrome (c), and linear invariant (d) bilinear interpolation between the four myocardial tensors at the vertices. Maps of tensor trace (e), FA (f), tensor mode (g), and tensor determinant (h) from the resultant interpolated tensor fields are shown for each interpolation method. The front left edge in the tensor glyph images corresponds to the bottom of the tensor invariant images.

was computed, and 1000 randomly resampled populations with replacement were constructed from the paired difference dataset. From each randomly resampled population the median was calculated. The 1000 median measures were sorted, and the asymmetric 95% confidence interval (CI) about the median was computed from the distribution for each tensor scalar measure. When this method is applied to paired angle differences between the primary eigenvectors or log-Euclidean tensor distances, only unsigned differences or distances can be computed. The median of the scalar tensor measure differences and the bootstrapped 95% CI of the median were compared to the zero-bias line (null hypothesis). If the paired differences are not significant, then the 95% CI will overlap with the zero-bias line. When the paired difference CI does not overlap with the zero-bias line, then the respective tensor interpolation method introduces a significant bias to the tensor field.

4   Results

Synthetic Example. Fig. 1 shows an example of bilinear interpolation between the four myocardial tensors at the vertices using the EU, LE, GL and LI methods. The tensors are visualized using superquadric glyphs [14]. Each edge represents a microstructural transformation that can be observed both histologically and with DT-MRI. EU only monotonically interpolates TR. EU negatively biases FA and positively biases DET (i.e. the so-called tensor swelling effect [9]). LE negatively biases TR and FA, and only monotonically interpolates DET. Both EU and LE heterogeneously bias MODE. GL monotonically and LI linearly interpolate all the tensor invariants including DET. In order to establish that monotonic or linear interpolation of the tensor invariants is the best interpolation method, we evaluated each tensor interpolation method using the experimentally measured DT-MRI datasets as follows.

Autocorrelation. The AC procedure resulted in AC lengths of 3 in all the directions for the rabbit heart data; 4 in the $x-$ and $y-$directions, and 2 in the $z-$direction for the porcine heart data; and 8 in the $x-$ and $y-$directions, and 6 in the $z-$direction for the human heart data. To ensure that the data was de-correlated, the data was decimated by the AC length in each dimension.

Bootstrap Statistics. Figure 2 shows that EU does not introduce a significant bias to TR nor DET, but it does negatively bias FA and positively bias MODE. AI and LE are nearly identical and show a negative bias for TR, FA, and DET;

(a) TR   (b) FA   (c) MODE   (d) DET   (e) 1st EV   (f) LE dist
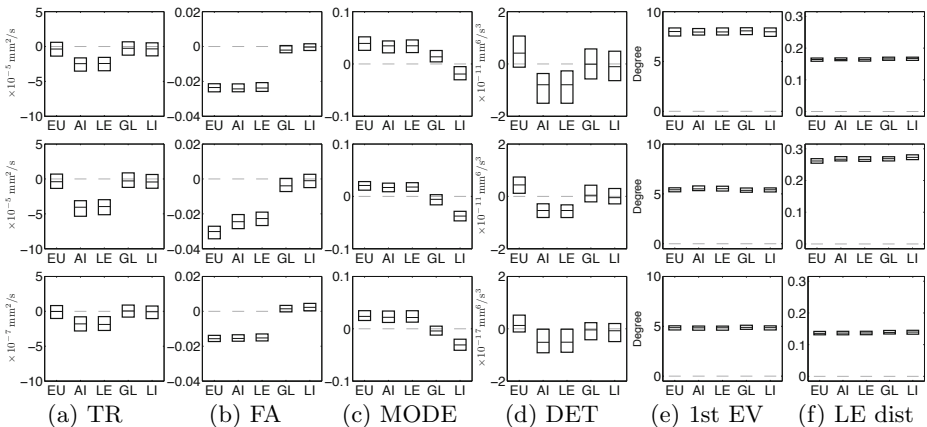
Fig. 2. Bootstrap statistics for tensor measures. The upper row corresponds to the rabbit heart data, the middle row to the porcine heart data, and the lower row to the human heart data. Results of the paired comparison for tensor trace (a), FA (b), tensor mode (c), tensor determinant (d), angle difference between primary eigenvectors (e), and log-Euclidean tensor distance (f) are shown for each interpolation method. Each black horizontal line represents the median of each measure, and each box represents the bootstrapped 95% confidence interval of the median. The light gray horizontal lines at zero represent the zero bias.

and a positive bias for MODE. GL shows no significant bias for TR, FA, MODE, nor DET. LI does not introduce a significant bias for TR, FA, nor DET, but it does negatively bias MODE. All of the tensor interpolation methods produce an equivalent and positive bias for the primary eigenvector and the log-Euclidean tensor distance metric.

5   Conclusion

The bootstrap comparison results demonstrate that GL and LI outperform EU, LE and AI in terms of tensor shape recovery. The bias introduced by AI and LE for recovery of TR and DET is a small ($\approx 2\%$) change relative to the absolute measures. Furthermore, current models do not make use of the TR information from the DT-MRI data because the conductivity tensor's eigenvalues have to be rescaled, hence this bias is not likely to be significant. The bias in FA introduced by EU, AI and LE is larger ($\approx 8\%$), and may significantly impact simulations when this data is incorporated into the computational model to rescale the conductivity tensor in regions of fibrosis and scar. Hence, accurate recovery of FA is important.

The magnitude of the bias in MODE by EU, AI, LE, and LI is similar ($\approx 4\%$). Only GL shows a distinct advantage as it interpolates MODE with no bias. Both electrophysiologic activation and mechanical tissue properties are known to be orthotropic, therefore accurate interpolation of MODE (lower bias) is likely beneficial.

For computational electrophysiology and mechanical modeling of the heart, orientation recovery is very important because the primary eigenvector (myofiber direction) strongly governs the direction of electrical activation and active contraction. All of the tested tensor interpolation methods introduce a $\approx 5° - 8°$ bias, which may introduce notable fiber "disarray" into computational models. A tensor interpolation method that better recovers tensor orientation is still needed.

The path interpolated by LI (respectively, GL) between two tensors lies in the 6-dimensional nonlinear manifold of tensors; this path has a projection onto the 3-manifold of tensor invariants (losing the directionality information). Here we interpolate on the 3-manifold, to linearly (respectively, monotonically) preserve the tensor shape attributes. The use of direct linear interpolation does not imply, nor is it motivated by, assumptions about global linearity, but naturally follows by considering the small neighborhood around a point to be homeomorphic to Euclidean space (valid for short distances), as given by the manifold structure. The paths are demonstrably close approximations of each other, and our fundamental ignorance of the true physical path on the tensor manifold makes it difficult to describe either LI or GL as "more meaningful." We therefore tested all the methods on real data.

In conclusion, if MODE recovery is important then GL should be used despite the computational cost. If MODE recovery is not critical then LI interpolation is an otherwise equivalent tensor interpolation method with reduced computational

cost, which is important when interpolating tensors to the coordinates of 5 to 25 million computational nodes found in whole heart electrophysiology models. EU, AI, and LE tensor interpolation have no distinct advantage for the interpolation of tensor shape and orientation information based on the comparisons presented herein.

References

1. Basser, P.J., Mattiello, J., Le Bihan, D.: Estimation of the Effective Self-Diffusion Tensor from the NMR Spin Echo. J. Mag. Res. B 103, 247–254 (1994)
2. Ennis, D.B., Kindlmann, G.: Orthogonal Tensor Invariants and the Analysis of Diffusion Tensor Magnetic Resonance Images. Mag. Res. Med. 55, 136–146 (2006)
3. Kindlmann, G., Ennis, D.B., Whitaker, R.T., Westin, C.F.: Diffusion Tensor Analysis with Invariant Gradients and Rotation Tangents. IEEE Trans. Med. Imaging 26(11), 1483–1499 (2007)
4. Kung, G.L., Nguyen, T.C., Itoh, A., Skare, S., Ingels Jr., N.B., Miller, D.C., Ennis, D.B.: The Presence of Two Local Myocardial Sheet Populations Confirmed by Diffusion Tensor MRI and Histological Validation. J Mag. Res. Imaging 34(5), 1080–1091 (2011)
5. Batchelor, P.G., Moakher, M., Atkinson, D., Calamante, F., Connelly, A.: A Rigorous Framework for Diffusion Tensor Calculus. Mag. Res. Med. 53(1), 221–225 (2005)
6. Fletcher, P.T., Joshi, S.: Riemannian Geometry for the Statistical Analysis of Diffusion Tensor Data. Signal Processing 87(2), 250–262 (2007)
7. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the Manifold of Multivariate Normal Distributions: Theory and Application to Diffusion Tensor MRI Processing. J. Math. Imaging Vis. 25(3), 423–444 (2006)
8. Pennec, X., Fillard, P., Ayache, N.: A Riemannian Framework for Tensor Computing. Int. J. Comp. Vis. 66(1), 41–66 (2006)
9. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean Metrics for Fast and Simple Calculus on Diffusion Tensors. Mag. Res. Med. 56, 411–421 (2006)
10. Kindlmann, G., San José Estépar, R., Niethammer, M., Haker, S., Westin, C.-F.: Geodesic-Loxodromes for Diffusion Tensor Interpolation and Difference Measurement. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 1–9. Springer, Heidelberg (2007)
11. Bi, C., Takahashi, S., Fujishiro, I.: Interpolating 3D Diffusion Tensors in 2D Planar Domain by Locating Degenerate Lines. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Chung, R., Hammoud, R., Hussain, M., Kar-Han, T., Crawfis, R., Thalmann, D., Kao, D., Avila, L. (eds.) ISVC 2010, Part I. LNCS, vol. 6453, pp. 328–337. Springer, Heidelberg (2010)
12. Yang, F., Zhu, Y.M., Magnin, I.E., Luo, J.H., Croisille, P., Kingsley, P.B.: Feature-Based Interpolation of Diffusion Tensor Fields and Application to Human Cardiac DT-MRI. Med. Image Anal. 16(2), 459–481 (2012)
13. Helm, P.A., Raimond, L.: Winslow at the Center for Cardiovascular Bioinformatics and Modeling and Dr. Elliot McVeigh at the National Institute of Health for provision of data
14. Ennis, D.B., Kindlmann, G., Rodriguez, I., Helm, P.A., McVeigh, E.R.: Visualization of Tensor Fields using Superquadric Glyphs. Mag. Res. Med. 53, 169–176 (2005)

Morphological Analysis of the Left Ventricular Endocardial Surface and Its Clinical Implications

Anirban Mukhopadhyay[1], Zhen Qian[2], Suchendra M. Bhandarkar[1], Tianming Liu[1],
Sarah Rinehart[2], and Szilard Voros[3]

[1] Department of Computer Science, The University of Georgia, Athens, GA, USA
[2] Piedmont Heart Institute, Atlanta, GA, USA
[3] Stony Brook University Medical Center, Stony Brook, NY, USA

Abstract. The complex morphological structure of the left ventricular endocardial surface and its relation to the severity of arterial stenosis has not yet been thoroughly investigated due to the limitations of conventional imaging techniques. By exploiting the recent developments in Multirow-Detector Computed Tomography (MDCT) scanner technology, the complex endocardial surface morphology of the left ventricle is studied and the cardiac segments affected by coronary arterial stenosis localized via analysis of Computed Tomography (CT) image data obtained from a 320-MDCT scanner. The non-rigid endocardial surface data is analyzed using an isometry-invariant *Bag-of-Words* (BOW) feature-based approach. The clinical significance of the analysis in identifying, localizing and quantifying the incidence and extent of coronary artery disease is investigated. Specifically, the association between the incidence and extent of coronary artery disease and the alterations in the endocardial surface morphology is studied. The results of the proposed approach on 15 normal data sets, and 12 abnormal data sets exhibiting coronary artery disease with varying levels of severity are presented. Based on the characterization of the endocardial surface morphology using the *Bag-of-Words* features, a neural network-based classifier is implemented to test the effectiveness of the proposed morphological analysis approach. Experiments performed on a strict leave-one-out basis are shown to exhibit a distinct pattern in terms of classification accuracy within the cardiac segments where the incidence of coronary arterial stenosis is localized.

Keywords: Ventricular endocardial surface, cardiovascular CT, non-rigid shape analysis, Bag-of-Words.

1 Introduction

The clinically observed relationship between the incidence and severity of Coronary Artery Disease (CAD) and the structural alterations in the left ventricular endocardial surface has not yet been formally studied due to inherent limitations of conventional cardiovascular imaging technologies. Since CAD is a leading cause of morbidity and mortality worldwide, techniques that improve diagnostic and prognostic effectiveness

have a potentially significant clinical impact. X-Ray Coronary Angiography (XRA) is an invasive technique that is a clinically accepted standard for assessment of vascular morphology and for quantifying the extent of vessel stenosis due to artherosclerotic plaque deposition. However, a comprehensive assessment of CAD necessitates a study of both, vascular morphology and cardiovascular function. Conventional myocardial functional assessment is based on a stress test that uses radionuclide or magnetic resonance (MR) perfusion imaging techniques. Since vascular morphology and cardiovascular function are imaged using separate modalities, the time and cost associated with a comprehensive assessment of CAD and the potential health risk to the patient associated with higher radiation doses are all significantly increased.

Computed Tomography Coronary Angiography (CTCA) is a non-invasive imaging technique that allows for robust quantification of vascular morphology and also has the potential for characterizing the atherosclerotic plaque composition [1]. When performed using a 320 Multirow-Detector Computed Tomography (MDCT) scanner, CTCA can yield images with an isotropic spatial resolution of 0.5 mm in a volumetric fashion. The resulting CTCA images, in addition to providing vascular morphology information, are capable of providing significant details about the endocardial surface structure, in particular, the structure of the *trabeculae* and *papillary muscles*.

We hypothesize that certain changes in the endocardial surface morphology bear a direct relationship to changes in cardiovascular function, i.e., the incidence and extent of stenosis in a specific coronary artery can be localized via analysis of morphological changes in the endocardial surface. The only known previous work along these lines is our previous work [11] which tackled this problem using two basic shape descriptors under the assumption that the endocardial surface could be treated as rigid, since all the images were taken at a relatively steady phase of the cardiac cycle. Although this work produced significant results, it had some inherent problems with regard to the selection of the shape descriptors. The two shape descriptors were proposed under the assumption of rigidity of the endocardial surface as observed in the MDCT images, i.e., the endocardial surface was *assumed* to be free of global deformation. Although the data was collected at 75% in the R-R cardiac cycle, i.e., at a relatively steady phase, the continuous motion of Left Ventricle (LV) demanded a more robust shape descriptor, i.e., one that is invariant to global deformation [9, 10]. In recent years, descriptors that are invariant to isometric deformations of an underlying surface have been studied in the context of shape-based retrieval in image databases. However, the problem in directly using existing shape descriptors developed for content-based image retrieval is that, these shape descriptors are designed with the goal of differentiating between two distinct classes of objects, e.g., "*humans*" with different poses versus "*dogs*" with different poses. In contrast, our goal is to differentiate between pathologies within the same class of objects, e.g., classify a particular LV segment as "*normal*" or "*diseased*", using its surface morphological properties.

Interest point feature-based methods have been used extensively in various computer vision algorithms [13] owing to the success of the *Scale Invariant Feature Transform* (SIFT) features [12] and *Speeded Up Robust Features* (SURF) [15]. The most advantageous aspect of interest point feature-based approaches is their treatment

of images as a collection of primitive elements, i.e., "*visual words*", and their natural progression to the use of well-developed methods from text search. The two prominent implementations of visual words as a backbone for shape description are *Shape Google-* a technique for shape-based search in large collections developed by Ovsjanikov *et.al.* [8] and, the work of Toldo *et.al.* [14]. By computing the frequency of the geometrical word occurrences in an image, a representation referred to as *Bag-of-Features* (BOF) is constructed for non-rigid shape description. As an extension to the BOF approach, we propose a novel shape analysis-based approach, termed as *Bag-of-Words* (BOW), to quantify the relationship between the incidence, severity and localization of CAD and the structural alterations in the LV endocardial surface.

Various computer vision applications have shown that visually similar images tend to share similar BOF descriptors. This property is useful for detection and description of similar images in a large-scale image database. The shape analysis community, on the other hand, has taken a long time to adopt BOF- or BOW-based approaches due to the lack of efficient and robust feature descriptors similar to SIFT. Some of the important properties of SIFT features include their inherent discriminative power combined with robustness to various image transformations. While several works in the research literature have proposed feature-based techniques for characterization of rigid shapes, very few are capable of dealing with non-rigid shape deformations [16]. To the best of our knowledge, this paper represents one of the first attempts, within the cardiovascular imaging community, to employ a BOW feature-based approach to compare non-rigid deformable shapes.
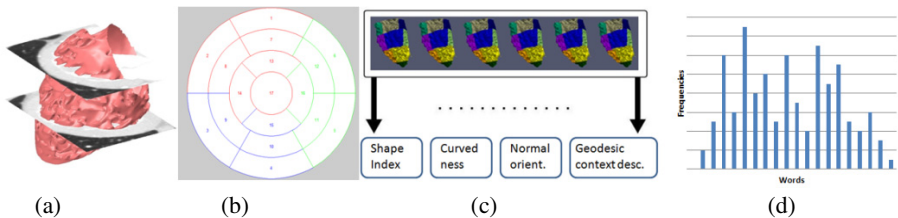
Fig. 1. Illustration of the sequence of steps in the morphological analysis of the LV endocardial surface: (a) accurate mesh segmentation, (b) generation of a 17-segment LV surface model with demarcation of coronary arterial territories (red: *LAD*, green: *LCX*, blue: *RCA*), (c) feature vector generation and (d) generation of the BOW histogram via vector quantization (*K*-means clustering).

We have proposed and implemented a BOW feature-based approach to encapsulate the local and global geometry as well as the local orientation information of the LV endocardial surface within a robust feature vector for the purpose of morphological analysis. The experimental results show successful localization of coronary arterial stenosis and thereby serve to strengthen the clinically observed relationship between the incidence and severity of CAD and alterations in the LV endocardial surface. The sequence of steps in the proposed approach for morphological analysis of the LV endocardial surface is depicted in Figure 1.

The remainder of the paper is organized as follows. In Section 2, the proposed LV surface segmentation and LV shape analysis procedures are detailed; in Section 3,

experimental results on the MDCT data sets are presented; in Section 4, the paper is concluded with a brief discussion of the proposed approach and an outline of directions for future work.

2   MDCT Image Segmentation and LV Shape Analysis

2.1   Left Ventricle Segmentation and Meshing

Three types of *trabeculae* morphologies are observed along the LV endocardial surface: (a) those that lie along the entire length of the ventricular wall forming prominent ridges; (b) those that are fixed at their extremities but free in the middle; and (c) those that connect the root of the papillary muscles and the ventricular wall. These *trabeculae* morphologies result in a complex endocardial surface topology. In order to adapt to the topological changes caused by the complex trabeculation structure, a 3D level set approach is employed to segment the LV endocardial surface. In order to suppress noise while still retaining the edges in the MDCT images, a median filter-based denoising procedure is employed on the 3D MDCT data prior to segmentation. The size of the median filter is empirically set to 7×7 based on the MDCT data set. A level set-based segmentation procedure without reinitialization, as proposed by Li *et al.* [5], is applied to the median-filtered 3D image data set followed by the marching cubes procedure [17] to generate the surface meshes. The surface meshes are subsequently denoised via a mean face normal filtering procedure proposed by Zhang and Hamza [7] to obtain the smooth shape of the myocardial surface of the left ventricle.

2.2   Data Preparation

In order to facilitate better understanding and localization of cardiac anatomy and pathology, the American Heart Association (AHA) has published recommendations for standardized myocardial segmentation [6]. We have adapted the AHA-approved 17-segment cardiac model [2] to divide the LV into 17 segments for more accurately localized shape analysis. The long axis of the LV is first computed, followed by division of the LV into 4 main parts, i.e., *apex, apical, mid-cavity* and *basal* along the longitudinal orientation. Division of the endocardial surface in the short axis view is tackled by exploiting knowledge of cardiac anatomy. Three landmark points are considered across the septum based on which the apical is divided into four parts and the mid-cavity and basal into six parts. Finally, the LV endocardial surface is divided into 17 segments.

2.3   Feature Description

500 surface points are randomly sampled from each of the 17 LV endocardial surface segments. Four types of surface descriptors are considered, inspired by Toldo et.al.

[14], to represent each sampled point. The first three descriptors are local whereas the fourth is a contextual descriptor. The descriptors are:

Shape Index (I_p)**:** The *shape index* I_p of a surface point *p*, as originally proposed by Koenderink [3] and subsequently modified by Zaharia and Preteux [4], is defined as a function of the two local principal surface curvatures k_p^1 and k_p^2 associated with surface point *p* and formulated as follows:
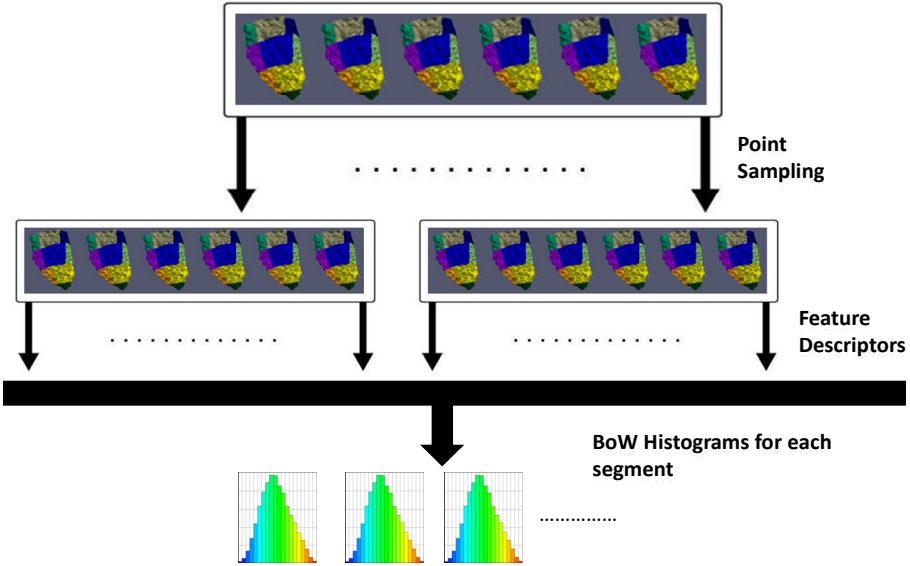
Fig. 2. Illustration of Bag-of-Words shape description and vector quantization

$$I_p = \frac{1}{2} - \frac{1}{\Pi} \arctan(\frac{k_p^1 + k_p^2}{k_p^1 - k_p^2})$$

Where, $k_p^1 > k_p^2$.

The range of the I_p values is [0, 1]. The value of I_p is not defined for planar surfaces. The shape index I_p provides a scale for representing basic elementary surface shapes such as *convex, concave, rut, ridge* and *saddle* [4]. The I_p value is invariant to scale and 3D rigid-body transformations (i.e., translation and rotation) in Euclidean space.

Curvedness(C_p): The *curvedness C_p* of a surface point p, as proposed by Koenderink [3], is defined as a function of the two local principal surface curvatures k_p^1 and k_p^2 associated with surface point p and formulated as follows:

$$C_p = \sqrt{(k_p^{1^2} + k_p^{2^2})/2}$$

where, $k_p^1 > k_p^2$.

Normal Orientation (θ_p)**:** The *normal orientation* θ_p at surface point p is defined as the angle between the unit normal vector at surface point p and the *XZ*-plane.

Geodesic Contextual Descriptor (GCD_p^{20})**:** The *geodesic contextual descriptor*, denoted by GCD_p^{20}, is a contextual descriptor that depends on the relative position of the surface point p vis-à-vis the other points on the surface segment. It is characterized by a 20-bin histogram which is generated by computing the normalized geodesic distance between the surface point p and 499 other sampled points on the surface segment. The geodesic contextual descriptor GCD_p^{20} is scale-invariant.

The feature extraction procedure described above results in a 23-tuple feature vector for each surface point p denoted by $F_p = (I_p, C_p, \theta_p, GCD_p^{20})$ as depicted in Figure 2.

2.4   Construction of Visual Vocabularies

The feature vectors F_p at each surface point are clustered in order to obtain the *visual words*. Assuming that the local descriptors are computed for a set of stable surface points, we quantize the feature vector space to obtain a compact representation for the vocabulary of *visual words*, in a manner similar to the Shape-Google approach [8]. A *vocabulary* is defined as a set of representative vectors in the descriptor/feature space, obtained by means of unsupervised learning, i.e., vector quantization via *k*-means clustering. More formally, a *vocabulary* is defined as a collection $V = \{ v_1, ..., v_k \}$ where v_i is the centroid of the ith cluster and the clusters represent the *visual words*. Here, we have chosen $k = 20$ in the *k*-means clustering algorithm for generating the final histogram.

3   Experimental Results

We employed the proposed methods for segmentation, meshing and endocardial surface shape description on 27 MDCT data sets consisting of 12 data sets from cardiac patients and 15 data sets from normal subjects. Incidence of single- or multi-vessel obstructive CAD was found in the three major coronary arteries using XRA, which was further confirmed by myocardial perfusion and fractional flow reserve tests performed on the patients.

The cardiac patients and normal subjects were subject to a contrast-enhanced CTCA scan on a 320-MDCT scanner using a standard CT angiography protocol with ECG gating. The resulting images were reconstructed at 75% in the R-R cardiac cycle to ensure minimal ventricular motion, so that the subsequent shape analysis is minimally affected by cardiac motion. The segmentation method described in Section 2.1 was used to generate topologically correct and geometrically accurate data.

3.1   Segmentation Results

The results of the LV segmentation have already been proven to be reasonably accurate [11]. Additionally, the spatial distribution of the trabeculation was observed to vary with the location within the LV; thus providing the rationale for using the 17-segment AHA model to perform localized shape analysis. Furthermore, the proposed segmentation approach has already shown a visually observable distinction in trabeculation between normal and diseased hearts, yielding classification accuracy greater than 80% with simple rigid shape descriptors and a nearest-neighbor classifier [11]. Our previous work has demonstrated the accuracy of the proposed segmentation method as well as its applicability for subsequent quantitative shape analysis.

3.2   Localization Results

A coronary artery is considered as *diseased or stenotic* if the extent of stenosis is 50% or greater. The LV myocardial segments are labeled as diseased by a cardiologist if they are supplied by stenotic arteries. An artificial neural network (ANN), employing a multilayer perceptron (MLP) architecture with a single hidden layer and a learning rate of 0.3, is used for the purpose of classification of the LV segments. The 20-bin histograms, generated via the vector quantization procedure (Section 2.3), for a particular LV segment from all the LV datasets are used as the inputs to the ANN. The classification procedure is carried out within a strictly leave-one-out setting. The output of the ANN is whether a particular LV segment can be classified as "*normal*" or "*diseased*".

The success rate for detection of stenosis in a specific coronary artery is shown in Figure 3. The classification results depict a clinically observed and intriguing relationship between the coronary arterial stenosis and the affected segment in the 17-segment AHA model. The lower classification accuracy in the basal area (segments 1-6) may be explained by the clinical observation that many of the coronary arterial stenoses in this study are located in the mid to distal portion of the coronary arteries that only effect the mid-cavity (segments 7-12) and apical portions (segments 13-16) of the LV endocardial surface. Furthermore, another probable reason for the lower classification rate in the basal area is that the apical and mid-cavity segments exhibit greater endocardial trabeculation structure than the basal segments, which translates to more reliable endocardial surface morphology information that can be used for the purpose of classification.
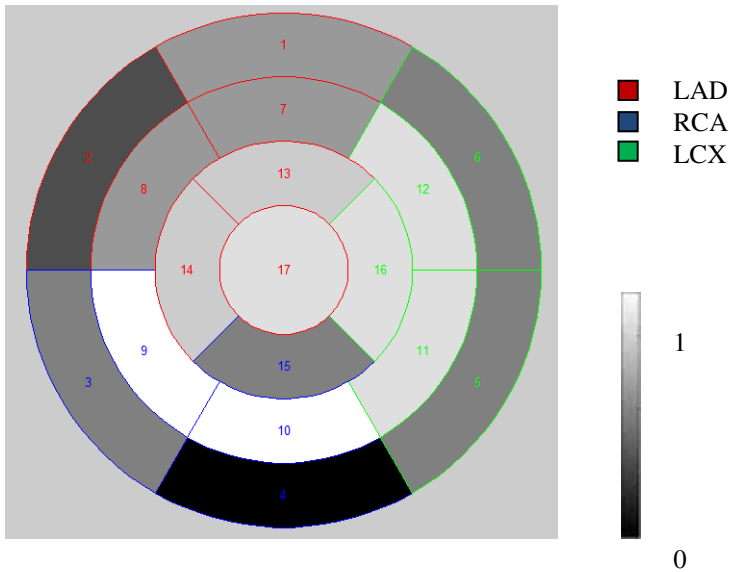
Fig. 3. Illustration of the classification accuracy for detection of coronary artery stenosis based on the change of surface morphology in the 17 LV segments in the AHA model. Higher gray values denotes higher classification accuracy and vice versa.

4 Discussion and Conclusions

To the best of our knowledge, this paper is one of the earliest works that studies the relationship between coronary artery stenosis and the morphological alterations in the LV endocardial surface using high-resolution MDCT data, and demonstrates its potential predictive value for the incidence and severity of CAD. This investigation also sheds new light on the localization of LV regions that are the most affected by coronary artery stenosis, a phenomenon which is yet to be fully explained. This association between the morphological features of the endocardial surface and cardiac functionality will be further explored in our future work. In particular, we aim to investigate the correlation between the endocardial surface morphology and the results of myocardial perfusion and fractional flow reserve tests in addition to the coronary arterial stenosis results obtained via XRA.

References

[1] Goo, S., Joshi, P., Sand, G., Gerneke, D., Taberner, A., Dollie, Q., LeGrice, I., Loiselle, D.: Trabeculae carneae as models of the ventricular walls: implications for the delivery of oxygen. Jour. Gen. Physiology 134(4), 339–350 (2009)
[2] Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., et al.: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. Circulation 105, 539–542 (2002)

[3] Koenderink, J.: Solid Shape. The MIT Press, Cambridge (1990)

[4] Zaharia, T., Preteux, F.: 3D Shape-based retrieval within the MPEG-7 framework. In: Proc. SPIE Conf. Nonlinear Image Proc. Pattern Anal. XII, vol. 4304, pp. 133–145 (2001)

[5] Li, C., Xu, C., Gui, C., Fox, M.D.: Level set evolution without re-initialization: a new variational formulation. In: Proc. IEEE Conf. CVPR, vol. 1, pp. 430–436 (2005)

[6] Medtronic Inc, The Visible Heart webpage,
    `http://www.visibleheart.com/index.shtml`

[7] Zhang, Y., Hamza, A.B.: Vertex-based diffusion for 3-D mesh denoising. IEEE Trans. Image Processing 16(4), 1036–1045 (2007)

[8] Ovsjanikov, M., Bronstein, A.M., Bronstein, M.M., Guibas, L.J.: Shape Google: a computer vision approach to invariant shape retrieval. In: Proc. NORDIA Workshop (2009)

[9] Sun, J., Ovsjanikov, M., Guibas, L.J.: A concise and provably informative multi-scale signature based on heat diffusion. In: Proc. Symposium on Geometry Processing (2009)

[10] Mémoli, F., Sapiro, G.: A theoretical and computational framework for isometry invariant recognition of point cloud data. Found. Comput. Math. 5(3), 313–347 (2005)

[11] Mukhopadhyay, A., Qian, Z., Bhandarkar, S., Liu, T., Voros, S.: Shape Analysis of the Left Ventricular Endocardial Surface and Its Application in Detecting Coronary Artery Disease. In: Metaxas, D.N., Axel, L. (eds.) FIMH 2011. LNCS, vol. 6666, pp. 275–283. Springer, Heidelberg (2011)

[12] Lowe, D.: Distinctive image features from scale-invariant keypoints. Intl. Jour. Computer Vision 60(2), 91–110 (2004)

[13] Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing 22(10), 761–767 (2004)

[14] Toldo, R., Castelllani, U., Fusiello, A.: The bag of words approach for retrieval and categorization of 3D objects. The Visual Computer 26(10), 1257–1268 (2010)

[15] Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. Computer Vision and Image Understanding (CVIU) 110(3), 346–359 (2008)

[16] Mitra, N.J., Guibas, L.J., Giesen, J., Pauly, M.: Probabilistic fingerprints for shapes. In: Proc. Symp. Geometry Processing, Vienna, Austria (2005)

[17] Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. Computer Graphics 21(4), 163–169 (1987)

Prior-Based Automatic Segmentation of the Carotid Artery Lumen in TOF MRA (PASCAL)

Jana Hutter[1,2], Hannes G. Hofmann[1], Robert Grimm[1], Andreas Greiser[3], Marc Saake[4], Joachim Hornegger[1,2], Arnd Dörfler[4], and Peter Schmitt[3]

[1] Pattern Recognition Lab, University Erlangen-Nuremberg, Germany
[2] School of Advanced Optical Technologies, Erlangen, Germany
[3] Siemens Healthcare, Magnetic Resonance, Erlangen, Germany
[4] Department of Neuroradiology, University Erlangen-Nuremberg, Germany

Abstract. In current clinical practice, examinations of the carotid artery bifurcation are commonly carried out with Computed Tomography Angiography (CTA) or contrast-enhanced Magnetic Resonance Angiography (ce MRA). Quantitative information about vessel morphology, extracted from segmentations, is promising for diagnosis of vessel pathologies. However, both above-mentioned techniques require the administration of contrast media. In contrary, non-ce MRA methods such as Time-of-Flight (TOF) provide fully non-invasive imaging without any exogenous contrast agent. The diagnostic value of TOF MRA, however, for assessment of the carotid bifurcation area can be hampered due to its susceptibility to irregular blood flow patterns. Conventional methods for lumen segmentation are very sensitive to such signal voids and produce inaccurate results. In this work, a novel, fully automatic 3D segmentation algorithm is proposed which uses prior knowledge about irregular flow patterns. The presented technique has been successfully tested on eleven volunteer datasets as well as in a patient case, offering the comparison to CTA images. The sensitivity could be increased by 29.2% to 85.6% compared to standard level set methods. The root mean squared error in diameter measurements was reduced from 4.85 mm to 1.44 mm.

1 Introduction

Stroke is one of the world leading causes of premature death [1] and shows increasing incidence numbers. Fast and reliable diagnosis of its cause is essential for treatment decisions. A common stroke source is stenosis in the internal carotid artery (ICA). Although CTA and Digital Subtraction Angiography (DSA) are still the gold standard for the carotid arteries, the risk for complications [2] and the exposition to ionizing radiation lead to a growing use of MRA. Besides the already widely used contrast-enhanced MR techniques, non-ce MR methods such as TOF allow contrast-agent-free, non-invasive imaging.

TOF relies on the inflow of blood perpendicular to the imaging plane to generate contrast between the vessel lumen and surrounding static tissue. Signal
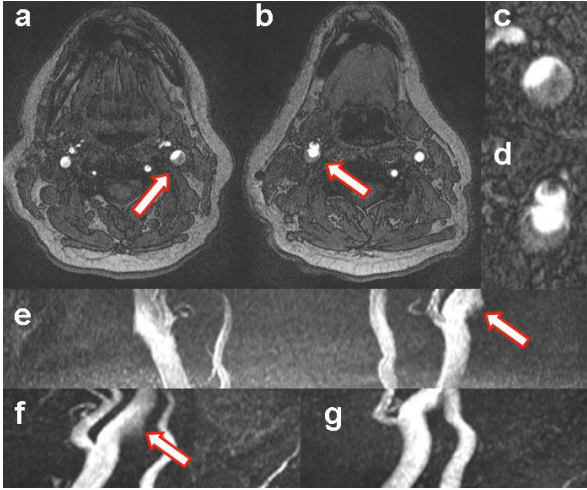
Fig. 1. Representative TOF dataset: (a) (b) Axial slices showing signal voids (indicated by arrows). (c) (d) Zoomed views. (e) Coronal MIP, (f),(g) sagittal MIPs.
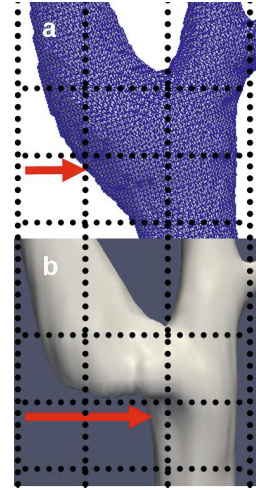
Fig. 2. (a) CTA and (b) MRA segmentation of the same patient

loss is observed in regions with irregular, fluctuating or reverse flow, which leads to intravoxel signal dephasing. In Fig. 1, typical TOF data of the carotids shows the characteristic artifacts pointed at by arrows. This susceptibility to irregular flow patterns [1] limits the usage for diagnosis as stenoses cannot be robustly identified. In critical regions, the lumen can show lower intensity close to the noise level, and strong signal gradients may occur inside. Though the true vessel delineation is still recognizable for the trained human eye, both effects constitute major problems for classical segmentation approaches as the sharp signal borders can easily be misinterpreted as the vessel wall. Especially the widely used Maximum Intensity Projection (MIP) visualization (Fig. 1(e)-(g)) suffers from the weak differentiation between the vessel signal and the surrounding signal.

Existing approaches can roughly be divided in stochastic and level-set or active contour based methods. There are algorithms specifically designed for non-ce MRA [3–8] as well as some focusing on the carotid arteries [3],[6], [9–12]. But only very few methods [3],[6] address the specific problem of visualizing the challenging bifurcation region in non-ce MRA.

In this work, we propose a method that accounts for the special properties of the MRA TOF acquisition and uses knowledge about regions prone to artifacts to identify even low intensity lumen regions. Contrary to most of the stated methods, our proposed method works fully automatically.

2   Segmentation

Comparison of the segmentations for both a CTA (Fig. 2(a), done with ITK-SNAP [13]) and a TOF (Fig. 2(b), done with a standard level set approach)
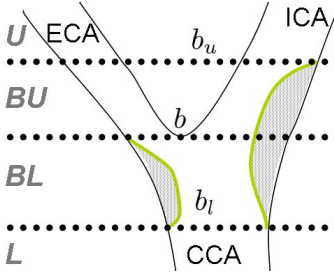
Fig. 3. Division of the ROI. Irregular flow zones are shaded
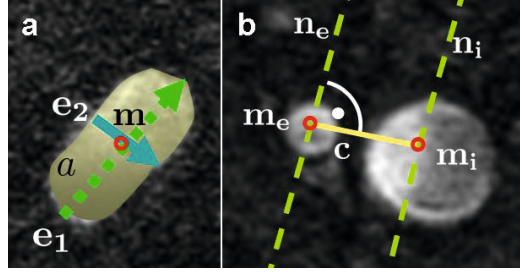
Fig. 4. (a) Ellipsoid fitting. (b) Detection of outer wall areas in **BU**.

scan reveals problems of conventional MRA TOF segmentation methods. Segmentation problems in the lower part of the ICA are indicated by arrows. Use of this segmentation for any quantitative evaluation would lead to significant underestimation of the vessel lumen. This problem is addressed by our fully automatic algorithm consisting of the following three steps.

2.1   MIP Generation and Threshold-Based Pre-Segmentation

In a first step, sagittal and transversal MIPs are generated from the slices of the 3D TOF scan. The regions of interest (ROI) are identified by adaptive thresholding. Noisy regions are eliminated as candidates by searching connected areas in slice direction. This leads to reduced computational effort as well as a more stable segmentation.

2.2   Detection of the Vessel Tree Skeleton and the Bifurcation

The common carotid artery (CCA) is differentiated from the vertebral artery in the first caudal slice based on the MIP pre-segmentation by size and connectivity to the bifurcation. The skeleton, consisting of center point localizations and bifurcation slices, is extracted as follows: The threshold result of the preceding slice $k-1$ is analyzed using ellipsoid fitting. This yields the major axis $\mathbf{e_1}_{k-1,v}$, the minor axis $\mathbf{e_2}_{k-1,v}$, the area $a_{k-1,v}$ and the center $\mathbf{m}_{k-1,v}$ for each detected vessel segment v, where $v = 0$ indicates the largest segment (see Fig. 4 (a)). Voxels on the major axis are used as seed points for the region-growing segmentation in slice k.

For each slice $f(k) = e(k) + a(k)$ is calculated and the bifurcation slice b is determined as the slice with maximal value of $f(k)$. The ellipsoid-similarity measure $e(k)$ favors elongated ellopsoids, like those which occur just under the bifurcation. The area ratio $a(k)$ helps to differentiate the main ICA-ECA bifurcation, where both vessels have generally close area values, from small ascending branches:

$$e(k) = \frac{\|\mathbf{e_1}_{k-1,0}\|}{\|\mathbf{e_2}_{k-1,0}\|} \qquad \text{and} \qquad a(k) = \frac{\min_v a_{k,v}}{\max_v a_{k,v}}. \tag{1}$$

The extracted information is used to find regions that are potentially sensitive to flow artifacts. Theoretical and computational fluid dynamic studies show that these irregular flow zones are located at the outer walls at the origins of ICA and ECA [14],[15] (Fig. 3, shaded regions).

Geometric Information. The volume is subdivided into four regions as illustrated in Fig. 3: The CCA part **L**, the lower bifurcation part **BL** where the CCA has a fully ellipsoidal form, **BU** above the bifurcation, and the upper part **U**. The distant parts **L** and **U** remain uncorrected, **BL** is fully corrected. Special attention is paid to **BU** where the region of the outer walls is critical. These regions are determined by identifying the outer halves of ICA and ECA. Therefore, their center points $\mathbf{m}_i$ and $\mathbf{m}_e$ (see Fig. 4(b)) are connected ($\mathbf{c} = \mathbf{m}_i - \mathbf{m}_e = (c_1, c_2)^{\mathrm{T}}$) and the normal direction is computed as $\mathbf{n} = (-c_2, c_1)^{\mathrm{T}} = (n_1, n_2)^{\mathrm{T}}$. Then the normal equations through the center points $n_i(x) : \mathbb{R} \mapsto \mathbb{R}$ and $n_e(x) : \mathbb{R} \mapsto \mathbb{R}$ are

$$n_i(x) = \frac{n_1}{n_2}x + m_{i,2} - \frac{n_1}{n_2}m_{i,1} \ , \ n_e \text{ accordingly.} \tag{2}$$

2.3   Level Set Evolution Using Prior Information

After these preprocessing steps, a full 3D level set evolution is applied. A contour of an open set ω is represented as the zero level of the higher dimensional scalar function $\Phi(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}$ with $\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}} \in \mathbb{R}^3$. $\Phi(\mathbf{x}) < 0$ holds for points inside the borders. The distance regularized level set evolution as proposed by Li et al. [16] is used, which enforces the level set function to keep the desired shape by the distance regularization term without the need for re-initializations. Also, the C^1 approximation of the Heaviside function $H_{app}(\Phi)$ and the approximated Dirac function $\delta_{app}(\Phi)$ are used.

The speed term $F(\Phi)$ consists of the regularization term $R(\Phi)$ and an external energy term $E(\Phi, \mathbf{x})$ pushing the level set evolution in the desired direction based on the image intensities $I(\mathbf{x})$:

$$F(\Phi) = \mu R(\Phi) + E(\Phi, \mathbf{x}). \tag{3}$$

A standard method to minimize the functional F is the steady state solution of the gradient flow which equals to

$$\frac{\partial \Phi}{\partial t} = -\mu \frac{\partial R}{\partial \Phi} - \frac{\partial E}{\partial \Phi}. \tag{4}$$

The external energy term $E(\Phi, \mathbf{x})$ is composed of the edge term $F_E(\Phi)$, the area term $F_A(\Phi)$ and the additional prior knowledge term $F_P(\Phi, \mathbf{x})$:

$$E(\Phi, \mathbf{x}) = \lambda F_E(\Phi) + \alpha F_A(\Phi) + \nu F_P(\Phi, \mathbf{x}), \tag{5}$$

where λ, α and $\nu > 0$ are the coefficients regulating the weight of the corresponding terms. The prior knowledge term $F_P(\Phi, \mathbf{x})$ consists of intensity deviations

from the mean background intensity level u_0, a weighting function $w(\mathbf{x})$ and the actual contour:

$$F_P(\varPhi, \mathbf{x}) = \int_\Omega P(\mathbf{x}) H_{app}(-\varPhi) d\mathbf{x} = \int_\Omega (I(\mathbf{x}) - u_0)^2 w(\mathbf{x}) H_{app}(-\varPhi) d\mathbf{x}. \qquad (6)$$

A novel key element in PASCAL is the weighting term $w(\mathbf{x})$ which depends on the regions as explained in Fig. 3 and Sect. 2.2. Contrary to approaches such as Scherl et al. [11], which uses integrated intensity deviations to differentiate plaque from lumen in CTA images, extracted skeleton information and physical knowledge about flow irregularities is included with $w(\mathbf{x})$:

$$w(\mathbf{x}) = \begin{cases} 0, & \text{if } x_3 < b_l, \\ 1, & \text{if } b_l \le x_3 < b, \\ \max(n_i(x_1) - x_2, 0) \cdot \max(x_2 - n_e(x_1), 0), & \text{if } b \le x_3 \le b_u, \\ 0, & \text{if } x_3 > b_u. \end{cases} \qquad (7)$$

The most important case is $b \le x_3 \le b_u$, which limits the regularization to the region close to the outer vessel wall just above the bifurcation by using the pre-calculated normals n_i and n_e. Using the variational principle and the relation $\frac{\partial H_{app}}{\partial \varPhi} = \delta_{app}$, the gradient descent flow equals to:

$$\frac{\partial \varPhi}{\partial t} = \mu \operatorname{div}(d_p(\|\nabla \varPhi\|^2) \nabla \varPhi) + \delta_{app}(\varPhi(\mathbf{x})) \left(\lambda \operatorname{div}(g \frac{\nabla \varPhi}{\|\nabla \varPhi\|^2}) + \alpha g + \nu P(\mathbf{x}) \right), \quad (8)$$

where $g = 1/(1 + \|\nabla I\|^2)$ and d_p as described in [16]. The level set equation is spatially discretized by using central differences for the spatial derivatives and temporally using finite differences with a time step Δt. The pre-segmentation serves as initial value $\varPhi^0(\mathbf{x})$.

3   Experiments and Results

3.1   Experimental Setup

TOF data was acquired in eleven volunteers (FOV $180 \, \text{mm} \times 180 \, \text{mm}$, imaging matrix 512×512, 3 slabs resulting in 51 slices, flip angle $25°$, slice thickness $0.5 \, \text{mm}$, TE=$3.76 \, \text{ms}$, TR=$23 \, \text{ms}$) on a clinical scanner at 3T (MAGNETOM Verio, Siemens Healthcare). Clinical data was acquired from a patient undergoing in addition a CTA examination (SOMATOM Definition AS+, Siemens Healthcare, rotation time $0.3 \, \text{s}$, increment $0.4 \, \text{mm}$, injection of $50 \, \text{mL}$ iodinated contrast agent, slice thickness $0.6 \, \text{mm}$, matrix size 512×512, 196 slices, in-plane resolution $0.46 \, \text{mm} \times 0.46 \, \text{mm}$).

Two segmentations have been performed. The standard level set (SLS) segmentation corresponds to the method proposed by Li et al. [16] with the following parameters: $\lambda = 20.0$, $\alpha = -10$, $\Delta t = 0.2$ and $\mu = 0.1$. The parameters for our proposed new method (PASCAL) have been identically chosen. The additional parameter ν was set to 0.08. The same threshold initialization and the same number of iterations were used for both methods.

3.2   Quantitative Evaluation

Segmentation Accuracy Study. For all twelve datasets, manual segmenta-
tions of a subset of slices were created by three independent readers including
an experienced radiologist and MR scientists familiar with TOF data as gold
standard. For the evaluation of the proposed segmentation method, sensitivity,
specificity, accuracy, the positive predictive value (PPV) and the Dice coefficient
were calculated both for the reference level set segmentation SLS and our novel
PASCAL method. Only detail patches around the vessel have been used for this
quantitative evaluation to ensure meaningful values especially for the specificity.
Furthermore, five people manually detected the bifurcation point in the TOF
image and the mean of the results was compared with the fully automatically
detected positions.

Vessel Diameter. A clinical experiment was carried out with the acquired
CTA and MRA patient dataset concentrating on quantitative measurements of
the bifurcation and ICA diameter, which is used to quantify stenoses, for example
with the NASCET criterion [2]. The diameter was measured at five positions in
the bifurcation (B_1-B_5) and three in the ICA (I_1-I_3) both in the CTA dataset
and the obtained MRA segmentation result.

3.3   Results

Segmentation results of the volunteer study are shown in Fig. 5 and Table 1,
comparing a standard level set approach and the PASCAL method. Our ap-
proach significantly outperforms the conventional segmentation. In particular in
the critical regions close to the bifurcation sensitivity values of 85.6% in com-
parison to 66.2% for the state of the art approach were achieved. The very high
specificity of over 99% as well as a stable high PPV valueof above 97% for both
approaches is particularly important to offer a reliable diagnosis of stenoses in
this region. The proposed algorithm was furthermore able to detect the skeleton
and the bifurcation region in all volunteer datasets. The root mean squared error
between the automatically and manually detected positions was 0.4 mm, which
is below the slice thickness of the used data. The quantitative measurements in
the clinical dataset (Table 2) yield diameters very close to the corresponding
CTA measurements. The root mean squared error of the diameter was reduced
from 4.85 mm to 1.44 mm.

Table 1. Evaluation of the segmentation results in comparison the gold standard

[%]	Bifurcation region (**BL**, **BU**)					Distant regions (**L**, **U**)				
	Sens.	Spec.	Acc.	Dice	PPV	Sens.	Spec.	Acc.	Dice	PPV
SLS	66.2	99.8	95.7	78.8	98.7	69.9	99.9	97.8	81.8	99.3
PASCAL	85.6	99.6	97.8	90.7	97.1	79.5	99.8	98.5	87.8	98.4
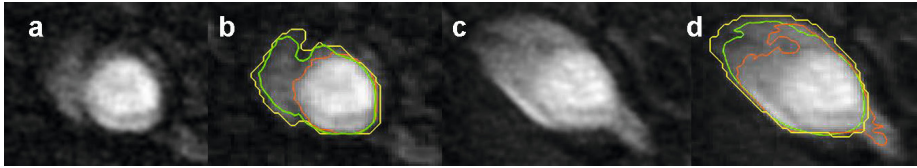
Fig. 5. (a) and (c) Two adjacent TOF slices. (b) and (d) Segmentation results. The yellow border demarcates the gold standard, the orange line the SLS and the green line shows the improved segmentation using PASCAL.

Table 2. Diameter of the carotis interna and the bifurcation in a clinical data set

		B_1	B_2	B_3	B_4	B_5	I_1	I_2	I_3
CTA SLS	[mm]	18.5	15.5	13.8	11.3	9.0	8.3	8.4	7.7
MRA SLS	[mm]	18.0	13.1	10.4	9.0	7.4	7.6	7.5	7.5
MRA PASCAL	[mm]	18.1	14.9	13.2	11.0	8.1	8.1	7.9	7.8

4 Discussion and Conclusions

A novel fully automatic segmentation approach for MRA TOF has been shown, especially adapted to the challenging drawback of this method: irregular flow patterns in dedicated regions. By including this knowledge and the segmented vessel skeleton into the level set formulation stable and reliable segmentation results in the carotid bifurcation region have been shown. The achieved segmentation result is an essential basis for simulation of hemodynamics as well as for quantitative measurements of stenosis degree or bifurcation angles which can be used as essential factors in computer aided diagnostics. With this algorithm, the use of standard TOF acquisitions, providing contrast-agent and ionizing radiation free imaging, in the clinical diagnosis and treatment decision workflow becomes feasible. Experimental data including the datasets and ground-truth segmentations are available online at http://www5.cs.fau.de/data/pascal.

References

1. Debrey, S.M., Yu, H., Lynch, J.K., Lövblad, K.-O., Wright, V.L., Janket, S.-J.D., Baird, A.E.: Diagnostic accuracy of magnetic resonance angiography for internal carotid artery disease: a systematic review and meta-analysis. Stroke 39, 2237–2248 (2008)
2. Michaely, H.J., Nael, K.: MRA of the Carotid Arteries. In: Parallel Imaging in Clinical MR Applications, pp. 291–306. Springer, Heidelberg (2007)

3. Frangi, A.F., Niessen, W.J., Hoogeveen, R.M., van Walsum, T., Viergever, M.A.: Quantitation of Vessel Morphology from 3D MRA. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 358–367. Springer, Heidelberg (1999)

4. Gao, X., Uchiyama, Y., Zhou, X., Hara, T., Asano, T., Fujita, H.: A fast and fully automatic method for cerebrovascular segmentation on time-of-flight (TOF) MRA image. J. Digit Imaging 24(4), 609–625 (2011)

5. Ladak, H.M., Milner, J.S., Steinman, D.A.: Rapid three-dimensional segmentation of the carotid bifurcation from serial MR images. J. Biomech. Eng. 122(1), 96–99 (2000)

6. Tang, H., van Onkelen, R.S., van Walsum, T., Hameeteman, R., Schaap, M., Tori, F.L., van den Bouwhuijsen, Q.J.A., Witteman, J.C.M., van der Lugt, A., van Vliet, L.J., Niessen, W.J.: A Semi-automatic Method for Segmentation of the Carotid Bifurcation and Bifurcation Angle Quantification on Black Blood MRA. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 97–104. Springer, Heidelberg (2010)

7. Watanabe, M., Kikinis, R., Westin, C.F.: Level set-based integration of segmentation and computational fluid dynamics for flow correction in phase contrast angiography. Acad. Radiol. 10(12), 1416–1423 (2003)

8. Yan, P., Kassim, A.A.: MRA Image Segmentation with Capillary Active Contour. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005, Part I. LNCS, vol. 3749, pp. 51–58. Springer, Heidelberg (2005)

9. van Bemmel, C.M., Viergever, M.A., Niessen, W.J.: Semiautomatic segmentation and stenosis quantification of 3D contrast-enhanced MR angiograms of the internal carotid artery. Magn. Reson. Med. 51(4), 753–760 (2004)

10. El-Baz, A., Farag, A.A., Gimel'farb, G., Hushek, S.G.: Automatic Cerebrovascular Segmentation by Accurate Probabilistic Modeling of TOF-MRA Images. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005, Part I. LNCS, vol. 3749, pp. 34–42. Springer, Heidelberg (2005)

11. Scherl, H., Hornegger, J., Prümmer, M., Lell, M.: Semi-automatic level-set based segmentation and stenosis quantification of the internal carotid artery in 3D CTA data sets. Med. Image Anal. 11(1), 21–34 (2007)

12. Suinesiaputra, A., de Koning, P.J.H., Zudilova-Seinstra, E., Reiber, J.H.C., van der Geest, R.J.: Automated quantification of carotid artery stenosis on contrast-enhanced MRA data using a deformable vascular tube model. Int. J. Cardiovasc. Imaging (2011)

13. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. Neuroimage 31(3), 1116–1128 (2006)

14. Perktold, K., Peter, R.O., Resch, M., Langs, G.: Pulsatile non-Newtonian blood flow in three-dimensional carotid bifurcation models: a numerical study of flow phenomena under different bifurcation angles. J. Biomed. Eng. 13(6), 507–515 (1991)

15. Marshall, I., Zhao, S., Papathanasopoulou, P., Hoskins, P., Xu, Y.: MRI and CFD studies of pulsatile flow in healthy and stenosed carotid bifurcation models. J. Biomech. 37(5), 679–687 (2004)

16. Li, C., Xu, C., Gui, C., Fox, M.D.: Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. 19(12), 3243–3254 (2010)

A Convex Relaxation Approach to Fat/Water Separation with Minimum Label Description

Abraam S. Soliman, Jing Yuan, James A. White,
Terry M. Peters, and Charles A. McKenzie

University of Western Ontario, London, ON, Canada
Robarts Research Institute, London, ON, Canada

Abstract. While Magnetic Resonance Imaging is capable of separating water and fat components in the body, mapping of magnetic field inhomogeneities is essential for the successful application of this process. In this study, we address the problem of field map estimation using a convex-relaxed max-flow method. We propose a novel two-stage approach that leads to the global optimum of the proposed problem. The first stage minimizes the signal residuals via a convex-relaxed minimum description length (MDL)-based approach. The MDL-based labeling model penalizes the total number of appearing labels, which helps to avoid field map errors when abrupt changes in field homogeneity exist. By exploring the whole range of possible frequency offsets, this stage ensures limiting the estimated field offset within certain boundaries where the global minimum resides. The second stage employs the output of the labeling model in a commonly used gradient-descent based method (known as IDEAL) to converge to the exact global minimum, i.e. the required value of the field offset. Experimental results for cardiac imaging, where challenging field inhomogeneities exist, showed that our method significantly outperforms over a widely-used technique for fat/water separation in terms of robustness and efficiency.

1   Introduction

The ability to separate fat from water in a magnetic resonance (MR) image is an important problem for a number of clinical applications. Bright fat signal can obscure underlying pathology and therefore suppression of the fat signal is required. In other cases, fat is considered an important diagnostic marker, and hence a clearer depiction of its signal, rather than its suppression, is desired. Common clinical applications of the latter include the diagnosis of non-alcoholic fatty liver diseases (NAFLD) [1], as well as a variety of bone marrow diseases [2]. However, expanding interest in the evaluation of myocardial fat infiltration and pericardial fat volume [3] justifies its optimization for cardiac imaging.

Among various available MR techniques, chemical-shift based (or Dixon-based) techniques have become the most commonly used methods to obtain a quantitative fat measurement [4]. Chemical-shift based methods are characterized by their unique ability to extract a fat-only image with positive contrast, unlike other techniques that either tend to suppress the fat signal, making the process of identifying fat voxels

ambiguous, or apply fat selective excitation which is sensitive to B0 and B1 inhomogeneities.

Unfortunately, a successful fat/water separation with Dixon-based techniques relies largely on the homogeneity of the magnetic field. In other words, the *mapping of the magnetic field inhomogeneities* – so called field map, cannot be decoupled from the fat/water separation process (see Sec 2.1). The field map estimation problem therefore, leads to a non-linear non-convex optimization problem, which has multiple local minima. An error in estimating the field map would propagate to the resultant water and fat images, causing what we term *fat/water swaps*. A "swap" is defined as assigning the main signal in a water-dominant voxel as fat, or vice-versa - an example of fat/water swaps is shown in Fig.1.
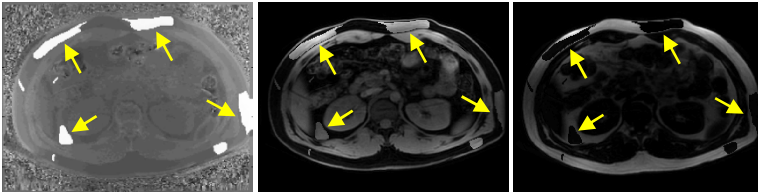
Fig. 1. Left to right: Fat/water swaps appearing in field map, water and fat components

A commonly-used technique in fat/water separation is the "Iterative Decomposition of water and fat with Echo Asymmetry and Least square estimation", abbreviated as IDEAL [5]. IDEAL is a Dixon-based method that acquires *at least three* echoes to estimate the field map, water and fat components. However, IDEAL is a local optimization method that heavily depends on the initialization process, and hence, convergence to the global optimum is not guaranteed. Moreover, it is a voxel-independent optimization, i.e. it does not enforce any global smoothness prior to the estimated field map. Yu *et al.* [6] proposed a region-growing technique[1] to address the flaws of IDEAL. This method implicitly imposes a spatial smoothness on the field map; however, it does not account for the abrupt changes in magnetic field that might exist at tissue/air interfaces, which might cause fat/water swaps. Although several techniques have been proposed in the literature [6-10] to address the field map estimation problem, only a few have been considered sufficiently robust for clinical use [8, 11]. Further, their application has been limited to 1.5 Tesla [11] where B0 field inhomogeneities are modest compared to higher field strengths.

In this work, we propose a novel field map estimation approach that can withstand abrupt changes in field homogeneity at higher field strengths, particularly at 3.0 Tesla, while guaranteeing smoothness of the estimated field map. Our method relies on prior knowledge of the periodic variation of signal residuals with the field map values [6-8]. We use a two-stage approach to reach the global minimum solution, and provide a high resolution mapping of the field inhomogeneities. First, a label-cost prior maxflow approach [12] is performed on the signal residues to converge *near* the global

---

[1] This nomenclature should not be confused with the conventional "region-growing" method used in general image processing applications.

optimum. The output is employed as an initial guess to the second stage, where a conventional gradient-descent IDEAL is applied to reach the exact field offset. Our method is tested for cardiac as well as abdominal images obtained at 3.0 Tesla, where challenging B0 field inhomogeneities commonly exist. Comparing to the region growing method [6], our approach has significantly improved the robustness of field map estimation process and has efficiently removed fat/water swaps.

2   Theory and Methodology

In the following sections, we first derive the signal equation to be minimized; then we introduce the multi-labeling convex relaxation model and its dual continuous max-flow formulation along with the minimum description length (MDL) principle, which are applied in the first stage of our approach. The proposed MDL-based labeling model penalizes the number of "appearing" labels, which helps to avoid the small regional fat/water swaps that might appear in the presence of severe and rapid changes of magnetic field. In other words, such MDL prior smooth out small-scale partitions, which usually correspond to regional fat/water swaps. The MDL-based labeling model results in a coarse estimation of the field map. This step guarantees a global minimization by labeling each pixel with a field map value located *near* its global optimum solution. The coarse estimate of field map serves as an initial guess for the second stage that consists of applying the IDEAL iterative process [5, 13]. A stopping criterion of < 1 Hz was used for the iterative process, in order to provide a field map with a sufficient resolution for clinical applications, particularly pericardial fat quantification. Once the final field map is obtained, water and fat components can be directly computed from Eq.2.

2.1   Signal Equation

Let $S(.)$ denote the signal acquired from a voxel v, containing a mixture of water and fat, such that:

$$S_v(t_n) = \left(\rho_{W,v} + \rho_{F,v} \cdot \sum_{m=1}^{M} \alpha_m \cdot e^{i2\pi\,\delta_m\,t_n} \right) \cdot e^{i2\pi\,\varphi_v\,t_n}, \tag{1}$$

where t_n denotes the echo-time (TE) shift ($n = 1, \ldots, N$) of the acquired signal; $\rho_{W,v}$ and $\rho_{F,v}$ are the water and fat components at voxel v, respectively; M is the number of fat peaks in the fat spectrum; δ_m is the frequency of the m-th peak with its corresponding amplitude α_m (Hz), such that $\sum_{m=1}^{M} \alpha_m = 1$; φ_v (Hz) is the local frequency offset at voxel v (i.e. the *field map*). We used a calibrated fat spectrum model as shown in [14], where $M = 6$ and the main fat peak is at ~ 420 Hz, relative to the water peak at 3.0 Tesla. Having three or more echo-times (TE) acquired (as described above), Eq.1 can be reformulated as follows:

$$S_v(t) = \Psi(\varphi_v, t) \cdot A(t) \cdot P_v, \tag{2}$$

where

$$\Psi(\varphi_v) = \begin{bmatrix} e^{i2\pi \varphi_v t_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{i2\pi \varphi_v t_N} \end{bmatrix}, \quad A = \begin{bmatrix} 1 & \sum_{m=1}^{M} \alpha_m \cdot e^{i2\pi \delta_m t_1} \\ \vdots & \vdots \\ 1 & \sum_{m=1}^{M} \alpha_m \cdot e^{i2\pi \delta_m t_N} \end{bmatrix},$$

$S_v(t) = [S_v(t_1), \dots S_v(t_n)]^T$, $P_v = [\rho_{W,v}, \rho_{F,v}]^T$. To estimate the required water and fat components $(\rho_{W,v}, \rho_{F,v})$, the frequency offset φ_v should be demodulated first. Hence, dropping the known echo-time shift (t_n), a non-linear least-squares cost function can be derived from (2) as follows:

$$\Gamma(\varphi_v) := \left\| A \cdot \widehat{P_v} - \widehat{\Psi}^{-1}(\varphi_v) \cdot S_v \right\|_2$$

$$\Gamma(\varphi_v) := \left\| (AA^\dagger - I) \; \widehat{\Psi}^{-1}(\varphi_v) \cdot S_v \right\|_2, \tag{3}$$

where $\widehat{P_v}$ and $\widehat{\Psi}(\varphi_v)$ are the estimated values of P_v and $\Psi(\varphi_v)$ respectively, I is the identity matrix, and † denotes the pseudo-inverse, $s.t. A^\dagger = [A^T A]^{-1} A^T$. However, two main problems are encountered when minimizing $\Gamma(\varphi_v)$: first, the non-convex property of the function, and second, it does not impose *a priori* smoothness on the estimated field map, as it is a voxel-by-voxel based strategy, and global minimization is not guaranteed. We address these problems by using a convex-relaxation approach that guarantees the global minimization, and implicitly includes the required smoothness of the field map.

It is important to note that, to maximize the signal-to-noise performance, the images are acquired at equally-spaced echo-time shifts (i.e. $TE_n - TE_{n-1} = \Delta TE$) [7]. In this case, $\Gamma(\varphi_v)$ is periodic with a period of $1/\Delta TE$ [6, 7]. This allows us to determine the lower and upper bounds, which are set to $[\pm 1/(2\Delta TE)]$, necessary for the max-flow model used in the first stage. The whole range is divided into ~20 equally-spaced values, which are used to label the input cost function, as described below.

2.2   A Continuous Max-Flow Approach to MDL-Based Potts Model

The Potts Model: In image processing, a multi-labeling problem assigns the optimal label $\varphi \in \varphi_1 \dots \varphi_L$ to each voxel. The Potts model is a labeling approach that minimizes the total perimeter of all one-label regions, without assuming any prior order for the labels. It results in a partition of the continuous domain Ω into L disjoint subdomains $\{\Omega_i\}_{i=1}^{L}$, as follows:

$$\min_{\{\Omega_i\}_{i=1}^{L}} \sum_{i=1}^{L} \int_{\Omega_i} \Gamma(\varphi_i, x) \, dx \; + \lambda \sum_{i=1}^{L} |\partial \Omega_i| \tag{4}$$

$$s.t. \bigcup_{i=1}^{L} \Omega_i = \Omega, \text{ and } \Omega_k \bigcap \Omega_m = \phi, \qquad \forall k \neq m$$

where $\Gamma(\varphi_i, x)$ is the cost of assigning label φ_i to location x, as defined in Eq.3, and $|\partial \Omega_i|$ measures the perimeter of each subdomain Ω_i, $i = 1, ..., L$.

The Potts model in Eq.4 can be efficiently solved by its convex-relaxation model as follows [15]:

$$\min_{u \in S} \left\{ E(u) := \sum_{i=1}^{L} \left(\int_{\Omega} u_i(x) \, \Gamma(\varphi_i, x) \, dx \ + \beta \int_{\Omega} |\nabla u_i| \, dx \right) \right\} \qquad (5)$$

where S is the convex constrained set of $u(x) := (u_1(x), ..., u_L(x))$:

$$S = \left\{ u(x) \mid \sum_{i=1}^{L} u_i(x) = 1, \ \ \forall x \in \Omega \, ; \ \ u_i(x) \in [0, 1], \ \ i = 1, ..., L \right\}.$$

Minimum Description Length (MDL)-Based Potts Model: The minimum description length (MDL) principle penalizes the number of appearances or labels in image labeling problems. It naturally leads to the use of fewer partitions or labels to describe the given image, without simultaneously over smoothing the underlying domain [12]. The MDL cost is introduced by adding the term γZ to the Potts model, where $Z = \#\{1 \leq i \leq L \mid \Omega_i \neq \emptyset\}$ gives the number of non-empty labels, i.e. a label-cost prior. Yuan *et al.* [12] showed that adding the label-cost prior to Eq.5 leads to the following convex-relaxed MDL-based Potts model, used here:

$$\min_{u \in S} \left\{ E(u) := \sum_{i=1}^{L} \left(\int_{\Omega} u_i(x) \, \Gamma(\varphi_i, x) \, dx \ + \beta \int_{\Omega} |\nabla u_i| \, dx \right) \right\} + \gamma \sum_{i=1}^{L} \max_{x \in \Omega} u_i(x) \qquad (6)$$

Fast Continuous Max-Flow Approach: The continuous max-flow approach [15, 16] to the MDL-based Potts model (Eq.6) used in this study is summarized below:

Let Ω be a continuous 2D image domain, L the number of labels, and Ω_i, $i = 1, ... L$ a copy of Ω assigned to the ith label. For each location $x \in \Omega$, a source flow $\rho_s(x)$ is streaming from the source node s to a labeled copy Ω_i, $s.t.$ $\forall \, \Omega_i$, $i = 1, ... L$, $\rho_s(x)$ is the same. Similarly for each $x \in \Omega$, a sink flow to the sink t is assigned. However, $\rho_i(x), i = 1, ... L$ may differ. A spatial flow $q_i(x) \, i = 1, ... L$ is also defined for each location. The continuous max-flow model can be formulated as follows:

$$\max_{\rho_s, \rho, q} \left\{ P(\rho_s, \rho, q) := \int_{\Omega} \rho_s \, dx \right\}, \qquad (7)$$

subject to the constraints:

   a. $(\mathrm{div} \, q_i - \rho_s + \rho_i - r_i \,)(x) = 0, \qquad i = 1, ..., L$

   b. $|q_i(x)| \leq C_i(x), \ \ \rho_i(x) \leq \Gamma(\varphi_i, x), \ \ \int_{\Omega} |r_i(x)| \, dx \leq \gamma \qquad i = 1, ..., L,$

where $C_i(x) = \beta$ is the capacity of the spatial flow $q_i(x)$, and $r_i(x)$ is the extra flow associated with the penalty of the number of labels. Yuan *et al.* [12] proved that the max-flow formulation (Eq.7) is dual to the convex-relaxed MDL-based Potts model, and results in an efficient flow maximization algorithm to Eq.6.

3   Experiments and Discussion

Our method was tested on 19 abdominal and cardiac images acquired on a 3.0 T MRI system (Discovery MR 750, GE Healthcare, Waukesha, WI). Cardiac images were acquired in different orientations (short-axis, 4-chambers and axial views) with a fast multi-echo GRE sequence using a 32-coil cardiac array [17]. Abdominal images were acquired with a 3D IDEAL-SPGR sequence [14] using an 8-coil array. Four and 6 equally-spaced echo-time shifts were used (interleaved acquisition for cardiac data). Each cardiac slice is acquired in one ~20s breath-hold while the 8 abdominal slices were acquired with parallel MRI acquisition (acceleration factor of 2) in one ~20s breath-hold. Matrix sizes of acquired images varied between 256x256 and 256x192.

The field map estimated from the first and final stages of our approach, as well as the fat and water components are shown below (Fig.2). For the abdominal example, TE/ΔTE = 1.04/0.828 msec and TR = 7.324 msec; for the cardiac example, TE/ΔTE = 2.28/1.54 msec and TR = 12.66 msec.
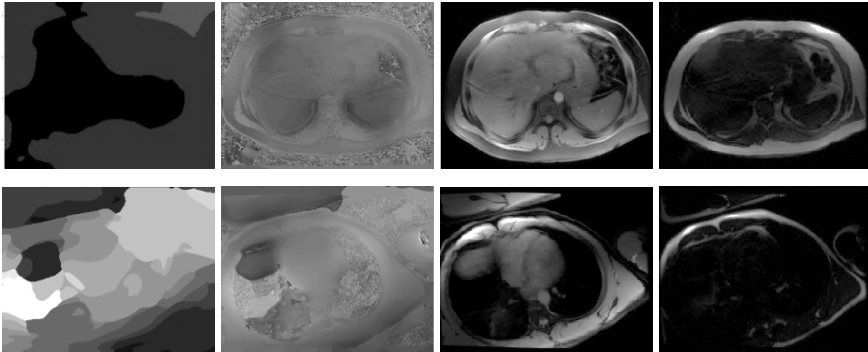
Fig. 2. Upper row: axial abdominal slice, lower row: 4-chambers cardiac view. Left to right: the coarse estimate of field map from the max-flow stage, final field map after the second stage, water and fat components.

Our results are compared to those obtained from the region-growing technique [6] on the same dataset, to show that fat/water swaps have been significantly reduced (Fig.3). A major drawback in the region-growing method is that it forces the field map smoothness using a 2D extrapolation approach, which does not account for the abrupt changes of field homogeneity, in particular in cardiac images. We have tackled this problem by employing a convex-relaxed labeling model that guarantees the global optimum. Moreover, it implicitly imposes smoothness on the estimated field map by penalizing small partitions, which correspond to regional fat/water swaps.
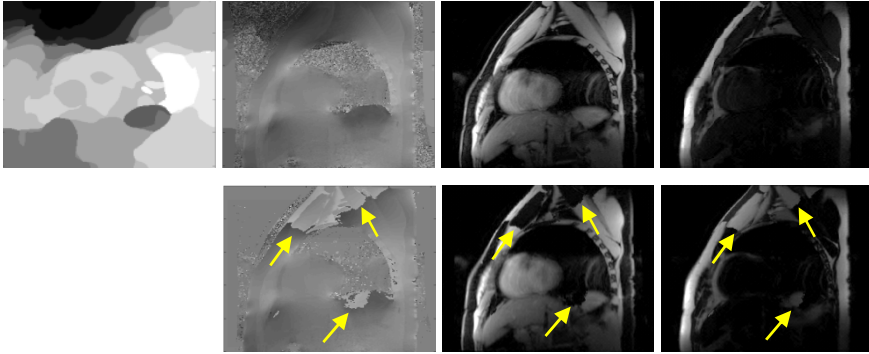
Fig. 3. Comparison between our approach (upper row) and the region-growing method (lower row) on a short-axis cardiac image. Yellow arrows indicate the locations of fat/water swaps that have clearly avoided by our method.

The processing time of our method is 1.5-5 minutes per image, depending on the level of smoothness and the underlying degree of inhomogeneities, compared to 6.5 minutes per image for the region-growing method. Moreover, the max-flow stage can be easily accelerated over modern computation frameworks, e.g. graphics processing units (GPU).

Taking into account the periodicity of $\Gamma(\varphi)$ [6, 7] (i.e. [$\pm$ 1/(2ΔTE)]), we found that using 20 equally-spaced field offsets for the first stage is sufficient to target the optimal label. However, relying only on the max-flow stage to reach the exact field offset would significantly increase the processing time, as we would use up to ~800 labels in order to achieve the same high resolution field map.

The robustness of field map estimation techniques is usually judged by the visual identification of fat/water swaps. However, we provide an approximate metric for the effectiveness of the technique by counting the number of pixels that have shown fat/water swaps in a user-defined region of interest – In abdominal images, the liver is assumed to be the region of interest, while in cardiac images the whole intra-thoracic space including the area surrounding the diaphragm is considered. An approximate average for the pixels showing fat/water swaps is ~274 $\pm$ 720 pixels with the region-growing technique vs. ~17 $\pm$ 32 pixels with our proposed method.

Comparing to a recent graph cut-based technique [8] currently used in clinical applications [11], our approach provides a field map with higher frequency resolution, with an estimation error less than 1 Hz over the whole image. This may have important utility for supporting accurate myocardial and pericardial fat quantification, given an expanding use of 3.0 Tesla field strengths for cardiac imaging.

References

1. Lall, C.G., Aisen, A.M., Bansal, N., Sandrasegaran, K.: Nonalcoholic fatty liver disease. Am J. Roentgeno. 190, 993–1002 (2008)
2. Haavardsholm, E.A., Bøyesen, P., Østergaard, M., Schildvold, A., Kvien, T.K.: Magnetic resonance imaging findings in 84 patients with early rheumatoid arthritis: bone marrow oedema predicts erosive progression. Ann. Rheum. Dis. 67, 794–800 (2008)
3. Iozzo, P.: Myocardial, Perivascular, and Epicardial Fat. Diabetes Care 34, S371 (2011)
4. Ma, J.: Dixon techniques for water and fat imaging. Magn. Reson. Imaging 28, 543–558 (2008)
5. Reeder, S., Pineda, A., Wen, Z., Shimakawa, A., Yu, H., Brittain, J., Gold, G., Beaulieu, C., Pelc, N.: Iterative decomposition of water and fat with echo asymmetry and least-squares estimation (IDEAL): Application with fast spin-echo imaging. Magn. Reson. Med. 54, 636–644 (2005)
6. Yu, H., Reeder, S., Shimakawa, A., Brittain, J., Pelc, N.: Field map estimation with a region growing scheme for iterative 3-point water-fat decomposition. Magn. Reson. Med. 54, 1032–1039 (2005)
7. Lu, W., Hargreaves, B.: Multiresolution field map estimation using golden section search for water-fat separation. Magn. Reson. Med. 60, 236–244 (2008)
8. Hernando, D., Kellman, P., Haldar, J., Liang, Z.: Robust water/fat separation in the presence of large field inhomogeneities using a graph cut algorithm. Magn. Reson. Med. 63, 79–90 (2010)
9. Soliman, A.S., Yuan, J., Vigen, K., White, J.A., Peters, T.M., McKenzie, C.A.: Robust Field Map Estimation using VARPRO and Multi-labeling Continuous Max-Flow. In: Proc. 20th ISMRM, p. 2508 (2012)
10. Soliman, A.S., Yuan, J., Vigen, K., White, J.A., Peters, T.M., McKenzie, C.A.: Fast Field Map Estimation withMulti-labeling Continuous Max-Flow. In: ISMRM Workshop on Fat-Water Separation (2012)
11. Kellman, P., Hernando, D., Arai, A.: Myocardial Fat Imaging. Current Cardiovascular Imaging Reports 3, 83–91 (2010)
12. Yuan, J., Bae, E., Boykov, Y., Tai, X.-C.: A Continuous Max-Flow Approach to Minimal Partitions with Label Cost Prior. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) SSVM 2011. LNCS, vol. 6667, pp. 279–290. Springer, Heidelberg (2012)
13. Reeder, S., Wen, Z., Yu, H., Pineda, A., Gold, G., Markl, M., Pelc, N.: Multicoil Dixon chemical species separation with an iterative least-squares estimation method. Magn. Reson. Med. 51, 35–45 (2004)
14. Yu, H., Shimakawa, A., McKenzie, C.A., Brodsky, E., Brittain, J.H., Reeder, S.B.: Multiecho water-fat separation and simultaneous R2* estimation with multifrequency fat spectrum modeling. Magn. Reson. Med. 60(5), 1122–1134 (2008)
15. Yuan, J., Bae, E., Tai, X.-C., Boykov, Y.: A Continuous Max-Flow Approach to Potts Model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 379–392. Springer, Heidelberg (2010)
16. Yuan, J., Bae, E., Tai, X.: A Study on Continuous Max-Flow and Min-Cut Approaches. In: Computer Vision and Pattern Recognition (CVPR), pp. 2217–2224 (2010)
17. Vigen, K.K., François, C.J., Yu, H., Shimakawa, A., Brittain, J.H., Reeder, S.B.: Multiecho IDEAL Cardiac Water-Fat Imaging. In: Proc. 17th ISMRM, p. 2775 (2009)

Regional Heart Motion Abnormality Detection via Multiview Fusion

Kumaradevan Punithakumar[1], Ismail Ben Ayed[1], Ali Islam[2],
Aashish Goela[3], and Shuo Li[1]

[1] GE Healthcare, London, Ontario, Canada
[2] St. Joseph's Health Care, London, Ontario, Canada
[3] London Health Sciences Centre, London, Ontario, Canada
{Kumaradevan.Punithakumar,Ismail.BenAyed,Shuo.Li}@ge.com,
Aashish.Goela@lhsc.on.ca, Ali.Islam@sjhc.london.on.ca

Abstract. This study investigates regional heart motion abnormality detection via multiview fusion in cine cardiac MR images. In contrast to previous methods which rely only on short-axis image sequences, the proposed approach exploits the information from several other long-axis image sequences, namely, 2-chamber, 3-chamber and 4-chamber MR images. Our analysis follows the standard issued by American Heart Association to identify 17 standardized left ventricular segments. The proposed method first computes an initial sequence of corresponding myocardial points using a nonrigid image registration algorithm within each sequence. Then, these points were mapped to 3D space and tracked using Unscented Kalman Filter (UKS). We propose a maximum likelihood based track-to-track fusion approach to combine UKS tracks from multiple image views. Finally, we use a Shannon's differential entropy of distributions of potential classifiers obtained from multiview fusion estimates, and a naive Bayes classifier algorithm to automatically detect abnormal functional regions of the myocardium. We proved the benefits of the proposed method by comparing the classification results with and without fusion over 480 regional myocardial segments obtained from 30 subjects. The evaluations in comparisons to the ground truth classifications by radiologists showed that the proposed fusion yielded an area-under-the-curve (AUC) of 95.9%, bringing a significant improvement of 3.8% in comparisons to previous methods that use only short-axis images.

1 Introduction

Accurate detection of motion abnormality of regional myocardial segments in MRI is essential in the diagnosis and treatment of coronary heart disease [7,11,15], the leading cause of death worldwide. The problem has attracted a recent research attention recently [7,11,15]. Unfortunately, existing MRI-based methods rely only on short-axis sequences [7,11,15]. However, the actual LV motion is a complicated combination of motions in 3D space. Little or no through-plane motion information is available from standard single view 2D sequences, which

severely limits the accuracy of 3D motion estimation. Therefore, exploiting information from other orthogonal image sequences can lead to a more accurate assessment of cardiac motion.

The purpose of our study is to develop a regional heart motion abnormality detection algorithm via multiview fusion, thereby exploiting the information from both short- and long-axis MRI sequences, namely, 2-chamber, 3-chamber and 4-chamber images. The proposed algorithm is based on a novel, multiview 3D motion estimation technique which consists of two main components, pre-processing and track-to-track fusion, both depicted in Fig. 1(a). Our 3D motion estimation is fundamentally different from existing 3D motion estimation methods, e.g., those based on incompressible models [2], deformable models [5,14], 3D harmonic phase [10] or short- and long-axis image registration [9], among others. Furthermore, it uses only standard clinical data (i.e., cine MRI[1]), unlike most of existing methods which either use data that are not available in regular clinical routine, such as displacement encoding with stimulated echoes (DENSE) MR images [14], or data that increase the scan time, such as myocardial tagging [5,10].

The proposed method first computes an initial sequence of corresponding myocardial points using a nonrigid image registration algorithm [4] within each 2D sequence, long- and short-axis, given a user-provided segmentation of the first frame. In order to provide a temporal smoothing to the dataset, we used a 3D extension of the nonlinear state transition model in [11]. Then, we propose to use an unscented Kalman smoother (UKS), a recursive nonlinear Bayesian approach, to obtain the state estimates and the corresponding covariance estimates. The state vector consists of position and velocity information of endo- and epi-cardial points over a cardiac cycle.

The main contributions of this study is a track-to-track multiview fusion based on a maximum-likelihood formulation which combines the UKS estimates from different views from short- and long-axis image sequences, thereby obtaining accurate 3D motion estimates. Track-to-track fusion problems are common in the multisensor fusion literature [1] but, to the best of our knowledge, were not investigated previously in medical imaging.

We prove the benefits of the proposed fusion in regional cardiac motion abnormality detection following a standard issued by the American Heart Association [3], and comparing the results with ground truth classifications by radiologists. The evaluations in comparisons to the ground truth classifications showed that the proposed fusion brings a significant improvement of 3.8% in area-under-the-curve (AUC) accuracy. The experimental analysis was carried over 480 regional myocardial segments obtained from 30 subjects (20 normal and 10 abnormal). We evaluated the classifier ability of Shannon's differential entropies (SDE) of normalized radial distance and endocardial segment volume *with and without fusion*. The classifier ability of these features were measured using receiver operating characteristic (ROC) curves with the corresponding AUCs, and the

---

[1] Cardiac cine MRI is the most widely used MR acquisition protocol in clinical routine due to its low processing time and complexity over other MR acquisition methods.
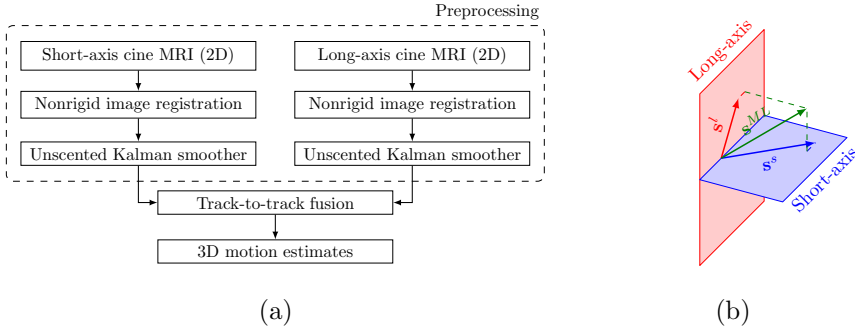
Fig. 1. (a) The proposed 3D motion estimation algorithm using 2D long and short-axis image fusion. (b) Illustration of a track-to-track maximum likelihood fusion $\mathbf{s}^{ML}$ using long- and short-axis state estimates $\mathbf{s}^l$ and $\mathbf{s}^s$.

Bhattacharyya distance metric [6]. We assessed the performance a via leave-one-subject-out approach. The proposed method yielded an AUC of 95.9%, whereas the method without multiview fusion (i.e. using only short-axis images as is the case in [11]) yielded an AUC of 92.1%.

2   Track-to-Track Fusion

Let $\mathbf{s}^s$ and $\mathbf{s}^l$ be the motion estimates obtained from, respectively, short- and long-axis images using UKS [12] and the registration algorithm [4]. The state vectors $\mathbf{s}^s, \mathbf{s}^l \in \{[\bar{\mathsf{x}}\ \mathsf{x}\ \dot{\mathsf{x}}\ \bar{\mathsf{y}}\ \mathsf{y}\ \dot{\mathsf{y}}\ \bar{\mathsf{z}}\ \mathsf{z}\ \dot{\mathsf{z}}\ \omega]^T\}$, where $(\mathsf{x}, \mathsf{y}, \mathsf{z})$ is a myocardial point on the 3D reference coordinate system corresponds to a pixel (i, j) in the image coordinate system. $(\mathsf{x}, \mathsf{y}, \mathsf{z})$ is computed using a transformation matrix constructed based on the information from DICOM header. $[\dot{\mathsf{x}}\ \dot{\mathsf{y}}\ \dot{\mathsf{z}}]$ is the velocity vector, $(\bar{\mathsf{x}}, \bar{\mathsf{y}}, \bar{\mathsf{z}})$ the mean position of $(\mathsf{x}, \mathsf{y}, \mathsf{z})$ over a cardiac cycle, and ω the angular frequency.

Having obtained state estimates $\mathbf{s}^s$ and $\mathbf{s}^l$ corresponding to the same tissue from different views, we now have to combine these estimates. In this study, we propose a Maximum Likelihood (ML) criterion (Refer to Fig. 1(b) for illustration) to compute a combined estimate $\mathbf{s}^{ML}$. We define the likelihood function as follows:

$$L(\mathbf{s}) = -\ln p(\mathbf{s}^s, \mathbf{s}^l|\mathbf{s})$$

$$\propto \left(\begin{bmatrix} \mathbf{s}^s \\ \mathbf{s}^l \end{bmatrix} - \begin{bmatrix} I \\ I \end{bmatrix} \mathbf{s}\right)^T P^{-1} \left(\begin{bmatrix} \mathbf{s}^s \\ \mathbf{s}^l \end{bmatrix} - \begin{bmatrix} I \\ I \end{bmatrix} \mathbf{s}\right) \qquad (1)$$

where

$$P = \begin{bmatrix} P^s & P^{sl} \\ P^{ls} & P^l \end{bmatrix} \qquad (2)$$

P^s and P^l are the covariances of $\mathbf{s}^s$ and $\mathbf{s}^l$, respectively, and P^{sl} is the cross-covariance between $\mathbf{s}^s$ and $\mathbf{s}^l$. I is an identity matrix. We compute the maximum

likelihood solution,

$$\mathbf{s}^{ML} = \arg\max_{\mathbf{s}} L(\mathbf{s}), \tag{3}$$

by solving $\nabla_{\mathbf{s}} L(\mathbf{s}) = 0$. This yields:

$$\mathbf{s}^{ML} = \left([I\ I]P^{-1}\begin{bmatrix} I \\ I \end{bmatrix}\right)^{-1}[I\ I]P^{-1}\begin{bmatrix} \mathbf{s}^s \\ \mathbf{s}^l \end{bmatrix} \tag{4}$$

Let $\mathcal{A} = P^s$, $\mathcal{B} = P^{sl}$ and $\mathcal{C} = P^l$. From inversion of a partitioned matrix, we have

$$P^{-1} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B}^T & \mathcal{C} \end{bmatrix}^{-1} = \begin{bmatrix} \mathcal{E} & \mathcal{F} \\ \mathcal{F}^T & \mathcal{G} \end{bmatrix} \tag{5}$$

where

$$\mathcal{E} = (\mathcal{A} - \mathcal{B}\mathcal{C}^{-1}\mathcal{B}^T)^{-1} \tag{6}$$
$$\mathcal{F} = -\mathcal{E}\mathcal{B}\mathcal{C}^{-1} \tag{7}$$
$$\mathcal{G} = \mathcal{C}^{-1} + \mathcal{C}^{-1}\mathcal{B}^T\mathcal{E}\mathcal{B}\mathcal{C}^{-1} \tag{8}$$

Substituting for P^{-1} in (4), we have

$$\mathbf{s}^{ML} = (\mathcal{E} + \mathcal{F}^T + \mathcal{F} + \mathcal{G})^{-1}(\mathcal{E} + \mathcal{F}^T)\mathbf{s}^s + (\mathcal{E} + \mathcal{F}^T + \mathcal{F} + \mathcal{G})^{-1}(\mathcal{F} + \mathcal{G})\mathbf{s}^l \tag{9}$$

Substituting for $\mathcal{E}$, $\mathcal{F}$ and $\mathcal{G}$ from (6)-(8) and applying matrix inversion lemma (refer to Appendix for derivation details), we get

$$\mathbf{s}^{ML} = (C - B^T)(\mathcal{A} + \mathcal{D} - \mathcal{B} - \mathcal{B}^T)^{-1}\mathbf{s}^s + (\mathcal{A} - \mathcal{B})(\mathcal{A} + \mathcal{C} - \mathcal{B} - \mathcal{B}^T)^{-1}\mathbf{s}^l \tag{10}$$

Substituting P^s, P^{sl} and P^l, we get

$$\mathbf{s}^{ML} = (P^l - P^{ls})(P^s + P^l - P^{sl} - P^{ls})^{-1}\mathbf{s}^s$$
$$+ (P^s - P^{sl})(P^s + P^l - P^{sl} - P^{ls})^{-1}\mathbf{s}^l \tag{11}$$

We assume that cross-covariance P^{sl}, P^{ls} between short- and long-axis observations are zeros. Thus, we have

$$\mathbf{s}^{ML} = P^l(P^s + P^l)^{-1}\mathbf{s}^s + P^s(P^s + P^l)^{-1}\mathbf{s}^l \tag{12}$$

3   Experiments

The data contains 30×3 short-axis image datasets (i.e., apical, mid-cavity and basal), each consisting of 20 functional 2D images acquired from 20 normal and 10 abnormal hearts. The data were acquired on 1.5T MRI scanners with fast imaging employing steady state acquisition (FIESTA) mode. In Fig. 2(a) and (b), we give a representative sample of the fusion results for end-diastolic and end-systolic phase of the cardiac cycle plotted against long-axis cine MR images. For each subject, three slices were respectively chosen from apical, mid-cavity

and basal frames. In Fig. 2(c), (d) and (e), we give a representative sample of the segmentation results for apical, mid-cavity and basal frames. The frames were automatically segmented following the standard issued by the AHA [3], given anatomical landmarks on the first frame. The results of 480 myocardial segments were compared with a single ground truth classification. We classify a segment as abnormal if that segment is hypokinetic, akinetic or dyskinetic.

We used two independent criteria to measure the performance of each classifier features, namely, the ROC curves with corresponding AUCs [8], and Bhattacharyya measure [6] to assess the discriminative power of each classifier features. Furthermore, we assessed the performance of the proposed approach via a *leave-one-subject-out* method.

(a) End-diastole (b) End-systole   (c) Apical   (d) Mid-cavity   (e) Basal
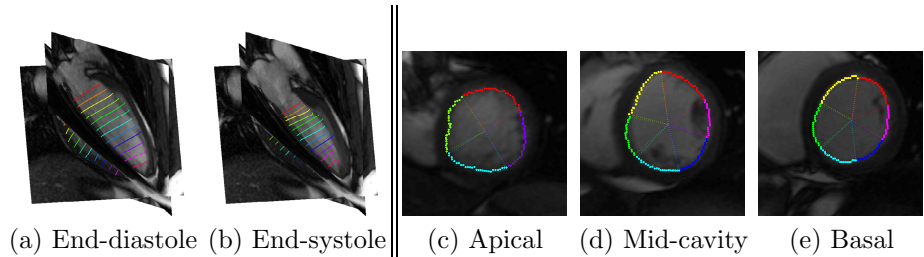
Fig. 2. (a) and (b): Representative examples showing the obtained fusion estimates plotted against long-axis MR images; (c), (d) and (e): Representative examples of segmented myocardium using the proposed approach. Apical, mid-cavity and basal frames were segmented, respectively, into 4, 6 and 6 segments following the standard in [3].

ROC, AUC and Bhattacharyya Measure: The ROC curves for classifier features SDEs of radial distance, segment area and segment volume are shown in Fig. 3. We used the same threshold for all segments and all slices. The ROC curves were obtained by varying such threshold. The AUCs corresponding to the ROC curves in Fig. 3 are reported in Table 1. The reported AUC values demonstrate that *multiview fusion* significantly improves the classifiers' ability in discriminating normal and abnormal heart motions.

We used the Bhattacharyya distance metric to evaluate the overlap between the distributions of classifier features over normal and abnormal motions. The Bhattacharyya metric [6] is given by $\mathbb{B} = \sqrt{1 - \sum_{y \in \mathbb{R}} \sqrt{f_N(y) f_A(y)}}$, where f_N and f_A are the distributions over, respectively, normal and abnormal motions. The higher $\mathbb{B}$, the lesser the overlap and, therefore, the better the discriminative ability of the classifier. The Bhattacharyya distance metrics reported in Table 1 demonstrate that the *multiview fusion* significantly improve the discriminative ability of the classifier features in detecting abnormal heart motion.

Classification Performance: The evaluations of classification performance in terms of accuracy, sensitivity and specificity are given by accuracy $= (T_P +$

Table 1. Comparison of the area under the ROC curve and Bhattacharyya distance metric for the methods with and without fusion (short-axis images only) [11].

	The proposed method (with multiview fusion)		Without fusion [11] (short-axis only)	
	AUC	Bhattacharyya	AUC	Bhattacharyya
SDE of segment volume/area	97.1	0.74	94.3	0.66
SDE of radial distance	95.9	0.70	92.1	0.61

(a)   (b)
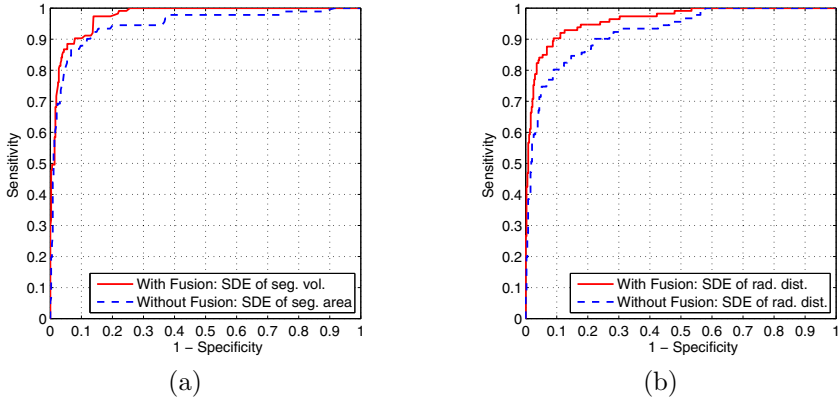
Fig. 3. Receiver operating characteristics of classifier features. The closer the curve to the left hand top corner, the better the classification performance.

$T_N)/(P + N)$, specificity $= T_N/N$, sensitivity $= T_P/P$, where T_P is true positives (number of segments correctly classified as *"Abnormal"*) and T_N true negatives (number of segments correctly classified as *"Normal"*). The number of *"Abnormal"* and *"Normal"* segments are P and N, respectively. Table 2 compares the classification performance of correctly classified hearts with the proposed method and the method that uses only short-axis images [11], using a *leaving-one-subject-out* method. In this approach, a naive Bayes classifier algorithm [13] is constructed from the SDEs of the segment area and normalized radial distance. Fig. 4 shows the quadratic decision boundary for normal/abnormal classification with the proposed method (with the UKS), where blue circles represent the normal function and red triangles the abnormal. The decision boundaries were constructed separately for apical, mid-cavity and basal slices learning from the remaining 29 subjects. The overall classification accuracy for the proposed method *with multiview fusion* is equal to 91.9%, with a sensitivity of 96.5% and specificity of 90.5%.

Table 2. The percentage of classification accuracy using a leaving-one-subject-out approach for the proposed track-to-track multiview fusion. The proposed method achieved an overall classification accuracy of 91.9%.

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Apex	90.8	96.9	88.6
Mid-cavity	95.0	95.3	94.9
Base	89.4	97.4	87.3
Overall	**91.9**	**96.5**	**90.5**

(a) Apical   (b) Mid-cavity   (c) Basal
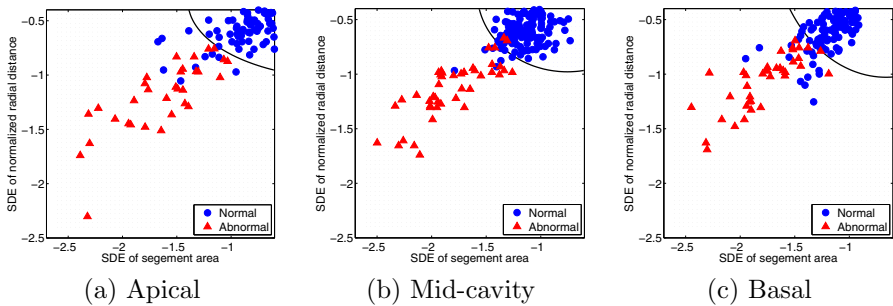
Fig. 4. Decision boundary for normal and abnormal regional myocardial functions using a Bayesian classifier

4   Conclusions

This study investigated a track-to-track multiview fusion approach to 3D LV motion estimation and regional abnormality detection. The proposed method uses several 2D cine MR image sequences, and yields state estimates in 3D space representing position and velocity information of myocardial points. A nonrigid image registration is used to obtain sequence of corresponding points and the UKS to track these points. Then, a track-to-track fusion method is proposed to combine UKS estimates from multiple images obtaining 3D state estimates. We show by an experimental evaluation that the proposed approach significantly improves the detection of regional abnormal motions in comparisons to previous approaches that use only the short-axis images.

References

1. Bar-Shalom, Y.: On the track-to-track correlation problem 26(2), 571–572 (April 1981)
2. Bistoquet, A., Oshinski, J., Skrinjar, O.: Left ventricular deformation recovery from cine MRI using an incompressible model. IEEE Trans. Med. Imag. 26(9), 1136–1153 (2007)

3. Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., Verani, M.: Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiologyof the American Heart Association. Circulation 105(4), 539–542 (2002)

4. Chen, H.-M., Goela, A., Garvin, G.J., Li, S.: A Parameterization of Deformation Fields for Diffeomorphic Image Registration and Its Application to Myocardial Delineation. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 340–348. Springer, Heidelberg (2010)

5. Chen, T., Wang, X., Chung, S., Metaxas, D., Axel, L.: Automated 3D motion tracking using gabor filter bank, robust point matching, and deformable models. IEEE Trans. Med. Imag. 29(1), 1–11 (2010)

6. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. 25(5), 564–577 (2003)

7. Garcia-Barnes, J., Gil, D., Badiella, L., Hernandez-Sabate, A., Carreras, F., Pujades, S., Marti, E.: A normalized framework for the design of feature spaces assessing the left ventricular function. IEEE Trans. Med. Imag. 29(3), 733–745 (2010)

8. Hanley, J., McNeil, B.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143(1), 29–36 (1982)

9. Lötjönen, J., Smutek, D., Kivistö, S., Lauerma, K.: Tracking Atria and Ventricles Simultaneously from Cardiac Short- and Long-Axis MR Images. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 467–474. Springer, Heidelberg (2003)

10. Pan, L., Prince, J., Lima, J., Osman, N.: Fast tracking of cardiac motion using 3D-HARP. IEEE Trans. Biomed. Eng. 52(8), 1425–1435 (2005)

11. Punithakumar, K., Ben Ayed, I., Islam, A., Ross, I.G., Li, S.: Regional Heart Motion Abnormality Detection via Information Measures and Unscented Kalman Filtering. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part I. LNCS, vol. 6361, pp. 409–417. Springer, Heidelberg (2010)

12. Sarkka, S.: Unscented Rauch–Tung–Striebel smoother. IEEE Trans. Automat. Contr. 53(3), 845–849 (2008)

13. Seber, G.A.F.: Multivariate Observations. John Wiley & Sons, Inc., Hoboken (1984)

14. Spottiswoode, B., Zhong, X., Lorenz, C., Mayosi, B., Meintjes, E., Epstein, F.: 3D myocardial tissue tracking with slice followed cine DENSE MRI. J. Magn. Reson. Imag. 27(5), 1019–1027 (2008)

15. Suinesiaputra, A., Frangi, A., Kaandorp, T., Lamb, H., Bax, J., Reiber, J., Lelieveldt, B.: Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. IEEE Trans. Med. Imag. 28(4), 595–607 (2009)

Global Assessment of Cardiac Function
Using Image Statistics in MRI

Mariam Afshin[1,3], Ismail Ben Ayed[2], Ali Islam[1], Aashish Goela[1],
Terry M. Peters[1,3], and Shuo Li[2]

[1] University of Western Ontario, London, Canada
[2] GE Healthcare, London, Canada
[3] Robarts Research Institute, London, Canada

Abstract. The cardiac ejection fraction (EF) depends on the volume variation of
the left ventricle (LV) cavity during a cardiac cycle, and is an essential measure
in the diagnosis of cardiovascular diseases. It is often estimated via manual seg-
mentation of several images in a cardiac sequence, which is prohibitively time
consuming, or via automatic segmentation, which is a challenging and computa-
tionally expensive task that may result in high estimation errors. In this study, we
propose to estimate the EF in real-time directly from image statistics using ma-
chine learning technique. From a simple user input in only one image, we build
for all the images in a subject dataset (200 images) a statistic based on the Bhat-
tacharyya coefficient of similarity between image distributions. We demonstrate
that these statistics are non-linearly related to the LV cavity areas and, therefore,
can be used to estimate the EF via an Artificial Neural Network (ANN) directly. A
comprehensive evaluation over 20 subjects demonstrated that the estimated EFs
correlate very well with those obtained from independent manual segmentations.

1 Introduction

One of the most important observations in diagnosing cardiovascular diseases, the car-
diac ejection fraction (EF), may decrease in the case of a heart attack or other problems
related to the heart valves or muscles. Furthermore, EF is an important indicator of
long-term prognosis for patients with coronary artery disease. Because the left ventricle
(LV) is the main pumping chamber of the heart, EF is usually measured using informa-
tion from the LV [7]. In routine clinical use, it is often estimated from several images in
a cardiac sequence using manual segmentation of the LV cavity, which is prohibitively
time consuming. While automatic LV segmentation can be used to compute the EF,
automatic LV segmentation is still acknowledged as a challenging, computationally ex-
pensive task, which has attracted impressive research attention in recent years. Existing
LV segmentation algorithms are based on traditional techniques, such as thresholding,
region-growing, edge detection and clustering [8,9,11], and energy minimization tech-
niques such as graph cuts [4, 13], active contours/level sets [2, 6], as well as active
appearance and shape models [1]. In general, segmentation algorithms require a care-
ful user initialization, intensive training, and a heavy computational load. Furthermore,
the ensuing segmentation results depend significantly on the choice of a set of *ad hoc*
parameters and training data, which may yield high errors in the computation of the

EF. These difficulties, inherent to segmentation algorithms, impede the automatic estimation of the EF in routine clinical use. A recent comprehensive review of cardiac image segmentation and its challenges can be found in [12]. While existing techniques are labor intensive, we believe that there are other characteristics of the images that can be computed with less effort, but that nevertheless correlate strongly with the EF. One such technique that we describe below is based on machine learning, which removes the need for image segmentation. From a simple user input in one single image, we build for all the images in a subject dataset (200 images) a statistic based on the Bhattacharyya coefficient [3] of similarity between image distributions. We demonstrate that these statistics are non-linearly related to the LV cavity areas (cf. Fig. 3) and therefore can be used to estimate the EF directly via an Artificial Neural Network (ANN). The proposed method consists of four main steps: (1) Image acquisition, (2) Building Image Statistics, (3) Applying Artificial Neural Network, and (4) Estimating Ejection Fraction. A comprehensive quantitative evaluation over 20 subjects demonstrates that the estimated EFs correlated very well with those obtained from manual segmentations.

2   Estimating Left Ventricle Volumes from Image Statistics

2.1   Building Image Statistics

Let $\mathscr{I}$ be a cardiac MRI sequence containing J frames[1], each comprising I slices[2], $\mathscr{I}_{i,j}$: $\Omega \subset \Re^2 \to \Re^+$ with $(i,j) \in [1\ldots I] \times [1\ldots J]$. To introduce our methodology for building an image statistic related to the LV cavity area for each image $\mathscr{I}_{i,j}$, $(i,j) \in [1\ldots I] \times [1\ldots J]$, let us consider the following definitions. (1) Let $\mathbf{I}$ be a reference image which we use for a simple user input (refer to the middle image in Fig. 1 b). For instance, in the experiments of this study, we used image $\mathscr{I}_{7,1}$ in each subject dataset. (2) Let $\Gamma_{in}, \Gamma_{out} : [0,1] \to \Omega$ denote two simple planar closed curves (e.g. squares) superimposed by the user on the reference image[3] $\mathbf{I}$ (refer to the middle image in Fig. 1 b), one placed within the cavity (the blue curve in Fig. 1 b) and the other enclosing the cavity (the red curve). Let us now superimpose systematically (without additional user effort) Γ_{out} onto each of the images in the subject dataset, as shown in Fig. 1. Then, we compute for each image a statistic based on the Bhattacharyya coefficient of similarity between image distributions (refer to Fig. 2), and demonstrate that *the obtained statistics are related to the areas of the LV*. Let $\mathbf{R}_\Gamma \subset \Omega$ be the region enclosed within Γ, $\Gamma \in \{\Gamma_{in}, \Gamma_{out}\}$, and $P_{\mathbf{R}_\Gamma, I}$ the kernel density estimate of the distribution of an image $I \in \mathscr{I}_{i,j}$, $(i,j) \in [1\ldots I] \times [1\ldots J]$, within region $\mathbf{R}_\Gamma$:

$$P_{\mathbf{R}_\Gamma, I}(z) = \frac{\int_{\mathbf{R}_\Gamma} K(z-I)dx}{a_{\mathbf{R}_\Gamma}}, \quad a_{\mathbf{R}_\Gamma} = \int_{\mathbf{R}_\Gamma} dx, \quad K(y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp^{-\frac{y^2}{2\sigma^2}} \quad (1)$$

where $a_{\mathbf{R}_\Gamma}$ is the area inside region $\mathbf{R}_\Gamma$ and K is the Gaussian kernel [10]. We consider the distribution of the image within the region enclosed by the blue curve in the

---

[1] The number of frames J is typically equal to 20 or 25.

[2] The number of slices I is typically equal to 10.

[3] The reference image is a mid-cavity slices at the end-diastolic time.

(a)　　　　　　　(b)
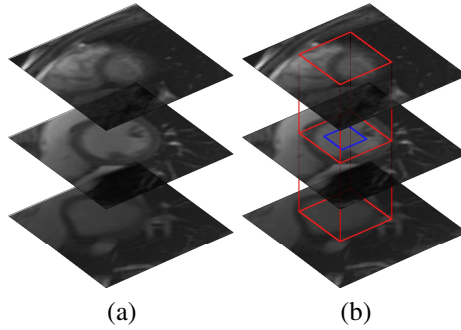
Fig. 1. (a) frame 1 (slices 1,7, and 10); (b) Γ_{in} (the blue curve within the cavity) and Γ_{out} (the red curve enclosing the cavity) are given by the user in the reference image (the middle image). Γ_{in} is used solely in the reference image to compute $P_{\mathbf{R}_{\Gamma_{in}},\mathbf{I}}$, whereas Γ_{out} is superimposed systematically (without additional user effort) to all the other images to compute $P_{\mathbf{R}_{\Gamma_{out}},\mathscr{I}_{i,j}}$.

reference image ($P_{\mathbf{R}_{\Gamma_{in}},\mathbf{I}}$) as an approximation of the distribution within the cavity, and the distribution of the region enclosed by the red curve in each image $\mathscr{I}_{i,j}$ ($P_{\mathbf{R}_{\Gamma_{out}},\mathscr{I}_{i,j}}$) as an approximation of the distribution of the entire left ventricle. Now consider the following measure of similarity between these two distributions in each image $\mathscr{I}_{i,j}$, $(i,j) \in [1 \ldots I] \times [1 \ldots J]$:

$$\beta^{i,j} = \mathbf{B}(P_{\mathbf{R}_{\Gamma_{in}},\mathbf{I}}, P_{\mathbf{R}_{\Gamma_{out}},\mathscr{I}_{i,j}}); \;\; \mathbf{B}(f,g) = \int_{R^+} \sqrt{fg}\,dz \qquad (2)$$

where the Bhattacharyya coefficient $\mathbf{B}(f,g)$ measures the amount of overlap (similarity) between two distributions f and g. The range of the Bhattacharyya coefficient is [0; 1], with 0 indicating no overlap between the distributions and 1 being a perfect match. The fixed [0; 1] range of the Bhattacharyya coefficient affords a conveniently practical appraisal of the similarity. More importantly, we expect the measure $\beta^{i,j}$ to be related to the cavity area in the corresponding image $\mathscr{I}_{i,j}$. This is demonstrated experimentally by the typical example in Fig. 2, and the corresponding variations of the cavity areas and the Bhattacharyya statistics in Fig. 3. *Note the strong similarity between the variations of the cavity areas and those of the Bhattacharyya statistics* (Fig. 3). Such similarity is reasonable since the more the distributions of the cavity and the LV overlap, the higher the cavity area.

2.2 Artificial Neural Network (ANN) Estimation of LV Cavity Areas

We constructed an Artificial Neural Network (ANN) to determine the nonlinear relation between the Bhattacharyya coefficients and the corresponding LV cavity areas (refer to Fig. 3 for an illustration of such non-linear relation). Following a back propagation ANN, a powerful machine learning technique [5], our feed-forward network consists of five layers, three hidden, one input, and one output (refer to Fig.4 (a) for an illustration). Let $P_{1,200}$ be the input of the network, a single row matrix containing the Bhattacharyya statistics, and let $T_{1,200}$ an output matrix containing the LV cavity areas:

I = $\mathscr{I}_{7,1}$ ... $\mathscr{I}_{7,7}$ (End-systolic) ... $\mathscr{I}_{7,19}$ (End-diastolic)

(a)   ...   (b)   ...   (c)

(d)   ...   (e)   ...   (f)

$\beta^{7,1} = 0.9188$ ... $\beta^{7,7} = 0.8754$ ... $\beta^{7,19} = 0.9609$
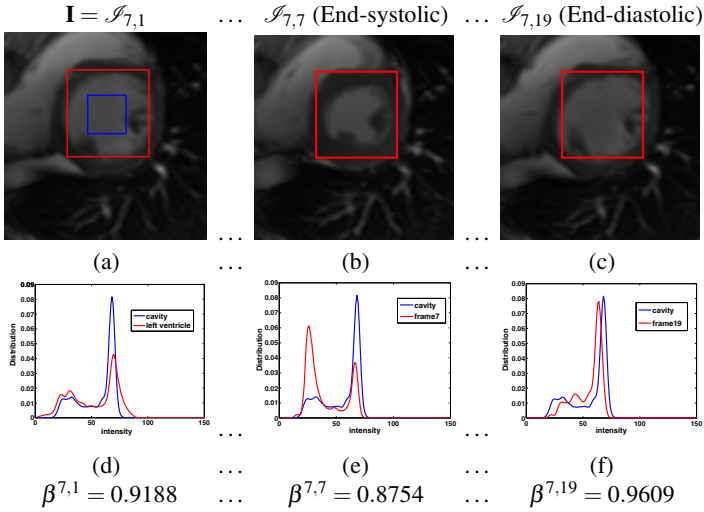
Fig. 2. Computing image statistics for the frames of slice 7 (middle slice): (a) reference image (red curve: Γ_{out}, blue curve: Γ_{in}); (b): frame 7 (end-systolic) and (c) frame 19 (end-diastolic); (d), (e), and (f) the corresponding distributions and Bhattacharyya measures (β^{ij}). *We observe that the variations of β^{ij} are similar to the variations of the LV cavity areas. For instance at the end of systole (the middle column), the smallest cavity area coincides with the lowest Bhattacharyya measure.*

(a)   (b)
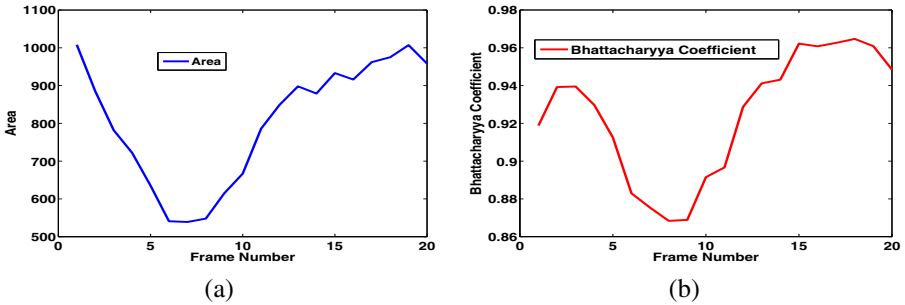
Fig. 3. (a) A typical example which shows the variations of LV cavity areas obtained from manual segmentations; (b) A typical example which shows the variations of the Bhattacharyya image statistics

$$P_{1,200} = [P^1, ..., P^j, ..., P^{10}] \quad with \quad P^j = \left[\beta^{1,j} ... \beta^{i,j} ... \beta^{20,j}\right] \tag{3}$$

$$T_{1,200} = [T^1, ..., T^j, ..., T^{10}] \quad with \quad T^j = \left[a^{1,j} ... a^{i,j} ... a^{20,j}\right] \tag{4}$$

To reduce the dimensionality of the inputs and outputs, we used principal component analysis (**PCA**) to transform 200 possibly correlated variables into a smaller set of uncorrelated variables (the first five components in our case):

$$IN_{1,5} = \mathbf{PCACOV}(P_{1,200}) \quad OUT_{1,5} = \mathbf{PCACOV}(T_{1,200}), \tag{5}$$

where **PCACOV** is the **PCA** transform function, $IN_{1,5}$ and $OUT_{1,5}$ denote the transferred inputs and outputs respectively.

The next step is to train the network using the transferred inputs ($IN_{1,5}$) and outputs ($OUT_{1,5}$). The network we built estimates the following non-linear mapping: $OUT_{1,5} = \mathbf{F}(IN_{1,5})$ where **F** is a nonlinear transfer function consisting of two hyperbolic tangent functions and a linear function, a common choice in the neural network literature [5]. As illustrated in Fig 4 (a), the resulting network consists of five layers, one input and one output containing 5 neurons each, both based on the linear function ($f(x) = x$), as well as three hidden layers containing 50, 25 and 50 neurons and based on the hyperbolic tangent, hyperbolic tangent and linear functions respectively. Let $INPUT_{5,19}$ and
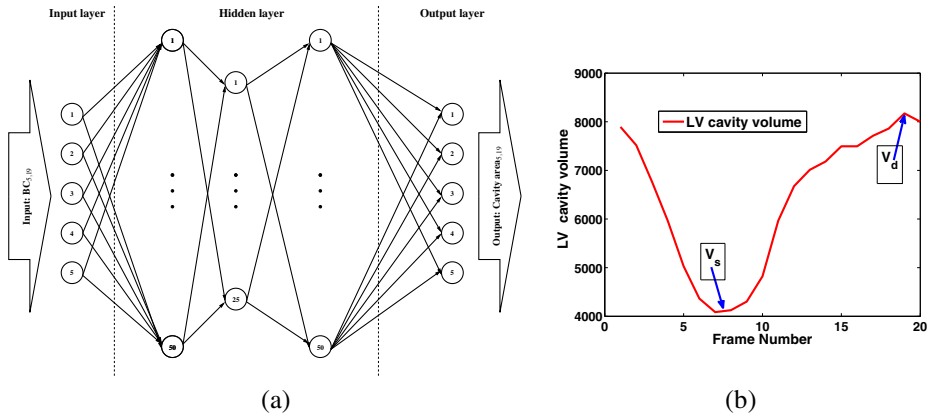
Fig. 4. (a) The ANN consists of one input layer, three hidden layers, and one output layer, (b) Variation of the volume of the LV cavity in each heart beat

$OUTPUT_{5,19}$ denote the training input and output of the neural network respectively:

$$INPUT_{5,19} = [(IN^1)^{-1}, ..., (IN^a)^{-1}, ..., (IN^{19})^{-1}] \quad st. \quad IN^a_{1,5} = \mathbf{PCACOV}(P_{1,200})$$
(6)

$$OUTPUT_{5,19} = [(OUT^1)^{-1}, ..., (OUT^a)^{-1}, ..., (OUT^{19})^{-1}]$$
(7)

$$st. \quad OUT^a_{1,5} = \mathbf{PCACOV}(T_{1,200})$$

To validate this procedure, *we employ a leave-one subject-out approach, where the test dataset was excluded from the training data*. For the current testing subject dataset, the LV cavity areas were estimated using the transferred subject Bhattacharyya statistics and the learned non-linear mapping **F** as depicted in Fig. 5 (b).

$$OUT_{test\,1,5} = \mathbf{F}(IN_{test\,1,5}) \quad T_{test\,1,200} = \mathbf{PCACOV}^{-1}(OUT_{test\,1,5})$$
(8)
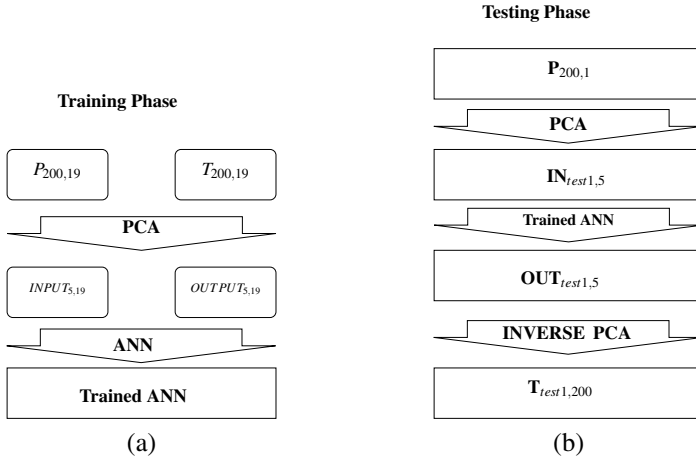
Fig. 5. (a) The training phase; (b) The testing phase: the estimated Bhattacharyya statistics are fed to the network and the corresponding LV cavity areas are predicted

2.3 Estimating the Cardiac Ejection Fraction from Image Statistics

Let V_s and V_d denote the smallest (end-systolic) and largest (end-diastolic) volumes of the LV in a cardiac cycle, respectively (Fig. 4(b)). The cardiac ejection fraction, EF, is given by: $EF = \frac{V_d - V_s}{V_d}$. The numerator measures the blood volume pumped by the left ventricle. We computed V_s and V_d by integrating the computed LV cavity areas in the sagittal direction.

3 Experimental Evaluations and Comparisons

A set of 2D short-axis cine magnetic resonance (MR) images of 20 subjects were acquired through the cardiac cycle on a 1.5T scanner with fast-imaging employing steady-state acquisition (FIESTA) image sequence mode. The acquisition parameters were as follows: TR=2.98 ms, TE=1.2 ms, flip angle=30 degree, and slice thickness=10 mm. Each subject's dataset consists of 20 frames throughout the cardiac cycle, each comprising 10 slices.

We used the proposed method to automatically compute the LV cavity areas, thereby estimating the LV cavity volumes and ejection fractions in each of the 20 subjects[4]. We proceeded to a *leave-one subject-out* validation approach, where the training used to compute the volumes of each subject is based on the other 19 subjects. The obtained volumes and ejection fractions were evaluated quantitatively by comparing them with those obtained from independent manual segmentation by an expert. Fig. 6 (a) depicts the computed LV cavity volumes for all 20 patients versus those obtained from the independent manual segmentations, as well as the identity line, which indicates an excellent correlation between manually and automatically computed volumes. In the next

---

[4] The dataset contains normal and abnormal cases.

(a)   (b)
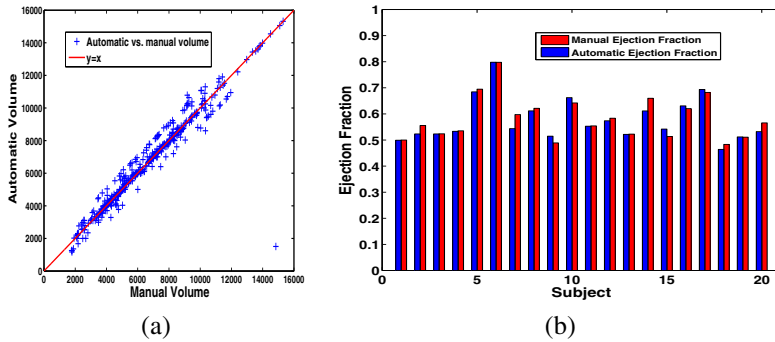
Fig. 6. (a) Automatic versus manual cavity volumes; (b) Automatic and manual EFs in 20 subjects

step, the estimated cavity volumes were used to estimate the EFs for all 20 subjects. Let EF_A be a vector containing the 20 automatically estimated EFs and EF_M a vector of the same size containing the EFs obtained form manual segmentations. Fig. 6 (b) depicts EF_A and EF_M, and confirms that the EFs computed with the proposed method are very close to those obtained from independent manual segmentations. We evaluated the conformity between the manually and automatically computed EFs (Table 1). First we evaluated the correlation coefficient, $CorrCoeff$, which measures the correlation between EF_A and EF_M. The range of $CorrCoeff$ is $[0,1]$, where 1 indicates a perfect fit between the vectors. The proposed method yielded a $CorrCoeff$ of 0.9635, which indicates a high conformity between manual and automatic ejection fractions. We then evaluated the mean and standard deviation of the norm of the difference between EF_A and EF_M: $Diff_{EF} = \|EF_A - EF_M\|$. The very low mean and standard deviation (std) of $Diff_{EF}$ (Table 1) indicates a high conformity between manual and automatic ejection fractions. We used a parametric test (two-tailed t-test) to estimate the conformity

Table 1. Statistical measures of the conformity between automatically and manually computed EFs and computation time (in seconds)

$CorrCoeff(EF_A, EF_M)$	$mean(Diff_{EF})$	$std(Diff_{EF})$	$CPU(s)$	$P-value(t-test)$
0.9635	0.0160	0.0163	0.2087	0.1778

between manually and automatically estimated ejection fractions that indicated the differences between EF_A and EF_M were not statistically significant ($P = 0.178$).

Figs 7 depicts automatically and manually computed volumes for three subjects. Fig. 7 (a) shows the best estimation in the 20 subjects, which corresponds to the lowest error, i.e., the lowest absolute difference between manually and automatically computed volumes. Fig. 7 (b) corresponds to the medium error (the medium estimation in the 20 subjects), and Fig. 7 (c) to the highest error (the worst estimation in the 20 subjects). The computation time is reported in Table 1. On a 2.2 GHz machine, a non-optimized MATLAB implementation took 0.2087 *seconds* to estimate the *EF per subject*.
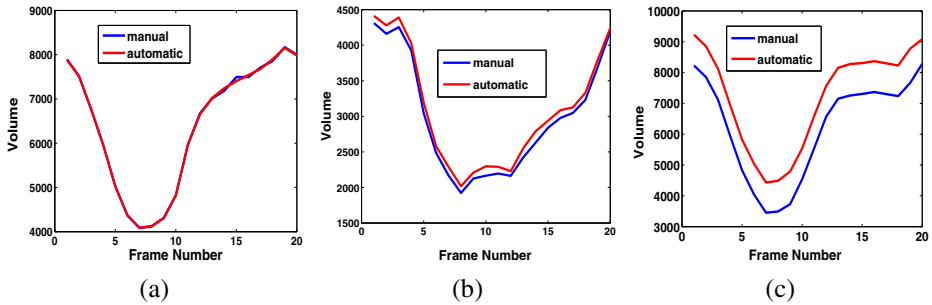
Fig. 7. Automatic versus manual cavity volumes: (a) the best case in the 20 subjects; (b) the medium case in the 20 subjects; (c) the worst case in the 20 subjects

4   Conclusion

This study investigated a real-time method for computing the cardiac EF directly (without segmentation) from image statistics via machine learning. These image statistics were based on the Bhattacharyya coefficients of similarity between image distributions, which were shown to be non-linearly related to the LV cavity areas. An ANN was used to find the relation between the image statistics and the corresponding LV cavity areas in each subject dataset. A comprehensive experimental evaluation over 20 subjects demonstrated an excellent conformity of the automatically estimated EFs to those computed from manual segmentations.

References

1. Andreopoulos, A., Tsotsos, J.K.: Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI. Medical Image Analysis 12(3), 335–357 (2008)
2. Ben Ayed, I., Li, S., Ross, I.: Embedding overlap priors in variational left ventricle tracking. IEEE Transaction on Medical Imaging 28(12), 1902–1913 (2009)
3. Ben Ayed, I., Li, S., Ross, I.G.: A statistical overlap prior for variational image segmentation. International Journal of Computer Vision 85(1), 115–132 (2009)
4. Ben Ayed, I., Punithakumar, K., Li, S., Islam, A., Chong, J.: Left Ventricle Segmentation via Graph Cut Distribution Matching. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 901–909. Springer, Heidelberg (2009)
5. Fausett, L. (ed.): Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Inc. (1994)
6. Fradkin, M., Ciofolo, C., Mory, B., Hautvast, G., Breeuwer, M.: Comprehensive Segmentation of Cine Cardiac MR Images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 178–185. Springer, Heidelberg (2008)
7. Han, H.C., Martin, R.P., Lerakis, G., Lerakis, S.: Prediction of the left ventricular ejection fraction improvement using echocardiography and mechanical modeling. Journal of the American Society of Echocardiography 18(7), 718–721 (2005)

8. Lee, H.Y., Codella, N., Cham, M., Weinsaft, J., Wang, Y.: Automatic left ventricle segmentation using iterative thresholding and an active contour model with adaptation on short-axis cardiac MRI. IEEE Transactions on Biomedical Engineering 57(4), 905–913 (2010)
9. Lynch, M., Ghita, O., Whelan, P.: Automatic segmentation of the left ventricle cavity and myocardium in MRI data. Computers in Biology and Medicine 36(4), 389–407 (2006)
10. Michailovich, O., Yogesh, R., Tannenbaum, A.: Image segmentation using active contours driven by the bhattacharyya gradient flow. IEEE Transaction on Image Processing 16(11), 2787–2801 (2007)
11. Pednekar, A., Muthupillai, R., Lenge, V.V., Kakadiaris, I.A., Flamm, S.D.: Automatic identification of the left ventricle in cardiac cine-MR images: Dual-contrast cluster analysis and scout-geometry approaches. Journal of Magnetic Resonance Imaging 23(5), 641–651 (2006)
12. Petitjean, C., Dacher, J.N.: A review of segmentation methods in short axis cardiac MR images. Medical Image Analysis 15(2), 169–184 (2011)
13. Zhu-Jacquot, J., Zabih, R.: Segmentation of the left ventricle in cardiac MR images using graph cuts with parametric shape priors. In: ICASSP, pp. 521–524 (2008)

Ultrasound and Fluoroscopic Images Fusion by Autonomous Ultrasound Probe Detection

Peter Mountney[1], Razvan Ionasec[1], Markus Kaizer[2], Sina Mamaghani[1], Wen Wu[1], Terrence Chen[1], Matthias John[2], Jan Boese[2], and Dorin Comaniciu[1]

[1] Siemens Corporate Research & Technology, Princeton, USA
[2] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. New minimal-invasive interventions such as transcatheter valve procedures exploit multiple imaging modalities to guide tools (fluoroscopy) and visualize soft tissue (transesophageal echocardiography (TEE)). Currently, these complementary modalities are visualized in separate coordinate systems and on separate monitors creating a challenging clinical workflow. This paper proposes a novel framework for fusing TEE and fluoroscopy by detecting the pose of the TEE probe in the fluoroscopic image. Probe pose detection is challenging in fluoroscopy and conventional computer vision techniques are not well suited. Current research requires manual initialization or the addition of fiducials. The main contribution of this paper is autonomous six DoF pose detection by combining discriminative learning techniques with a fast binary template library. The pose estimation problem is reformulated to incrementally detect pose parameters by exploiting natural invariances in the image. The theoretical contribution of this paper is validated on synthetic, phantom and *in vivo* data. The practical application of this technique is supported by accurate results (< 5 mm in-plane error) and computation time of 0.5s.

1 Introduction

Percutaneous and minimally-invasive cardiac procedures are progressively replacing conventional open-heart surgery for the treatment of structural and rhythmological heart disease. Catheters are used to access target anatomy through small vascular access ports. This greatly reduces recovery time and the risk of complications associated with open surgery. Without direct access and visualization, the entire procedure is performed under imaging guidance. There are two established modalities currently used in operating rooms to provide real-time intra-operative images: X-ray fluoroscopy (Fluoro) and transesophageal echocardiography (TEE). Fluoro provides high quality visualization of instruments and devices, which are typically radiopaque, while TEE and more recently 3D TEE can image soft-tissue with great detail. Nevertheless, the complementary nature of TEE and Fluoro is barely exploited in today's practice where the real-time acquisitions are not synchronized and images are visualized separately in misaligned coordinate systems.

Recently, the fusion of Fluoro and TEE has been proposed using either hardware or image based methods. Hardware based approaches [1],[2] attach additional devices to

the ultrasound probe such as electromagnetic [1] or mechanical [2] trackers and align the device and Fluoro coordinates systems through calibration. Image based methods [3],[4],[5] attempt to use the appearance of the TEE probe in the Fluoro image to estimate the pose of the probe in the Fluoro coordinate system. These methods are attractive because they do not require the introduction of additional equipment into the theatre which may disrupt clinical workflow.

Image based pose estimation is well studied and the problem may be considered solved when the correspondence between 2D image points and a 3D model are known. Unfortunately, the appearance of the TEE probe in the Fluoro image makes establishing the correspondence challenging. The probe's appearance lacks texture or clear feature points and can be homogenous under low dose or close to dense tissue. To alleviate this problem, markers [5] may be retro fitted to the TEE probe. The pose of the probe is estimated using well established computer vision techniques, however, the addition of markers increases the overall size of the probe. Alternatively the natural geometry of the probe may be used to estimate its pose [3],[4]. The authors use a 2D/3D registration technique to refine the probe's pose estimation and optimal results are obtained using two biplane images. The method is robust for small pose changes (10 mm / 10°), however, it requires manual initialization and does not update the registration in real-time, both of which are important in the clinical setting.

In the paper we propose a robust and fast learning-based method for the automated detection of the TEE probe pose, with six degrees of freedom, from Fluoro images. A probabilistic model-based approach is employed to estimate candidates for the in-plane probe position, orientation and scale parameters. Digitally reconstructed radiography (DRR) in combination with a binary template library is introduced for the estimation of out-of-plane rotation parameters (pitch and roll). The approach does not require manual initialization, is robust over the entire pose parameter space, and independent of specific TEE probe design / manufacturer. The performance of the algorithm is demonstrated on a comprehensive dataset of *in vivo* Fluoro sequences and validated on simulated and phantom data.

2   Fusion Framework

Information from a TEE volume can be visualized in a Fluoro image by aligning the TEE and C-arm Fluoro coordinate systems. A point in the ultrasound volume Q^{TEE} can be visualized in the Fluoro image at coordinate Q^{Fluoro} using the following transformation

$$Q^{FluoroImage} = P_{projection} R_{xz} T_d R_\gamma R_\alpha (R_{TEE}^W Q^{TEE} + T_{TEE}^W)$$

where $P_{projection}$ is the projection matrix, R_{xz} and T_d are the transformation from detector to world coordinate system, R_γ and R_α are the angulations of the C-arm and R_{TEE}^W and T_{TEE}^W are the rotation and position of the TEE probe in the world coordinate system such that $R_{TEE}^W = R_\alpha^{-1} R_\gamma^{-1} R_{xz}^{-1} R_{TEE}^{FluoroDetector}$ and $T_{TEE}^W = R_\alpha^{-1} R_\gamma^{-1} T_d^{-1} R_{xz}^{-1} R_{TEE}^{FluoroDetector}$.

The TEE volume and Fluoro image can be aligned if position $T_{TEE}^{FluoroDetector} = (x, y, z)$ and orientation $R_{TEE}^{FluoroDetector} = (\theta r, \theta p, \theta y)$ of the TEE probe are known in the Fluoro detector coordinate system.
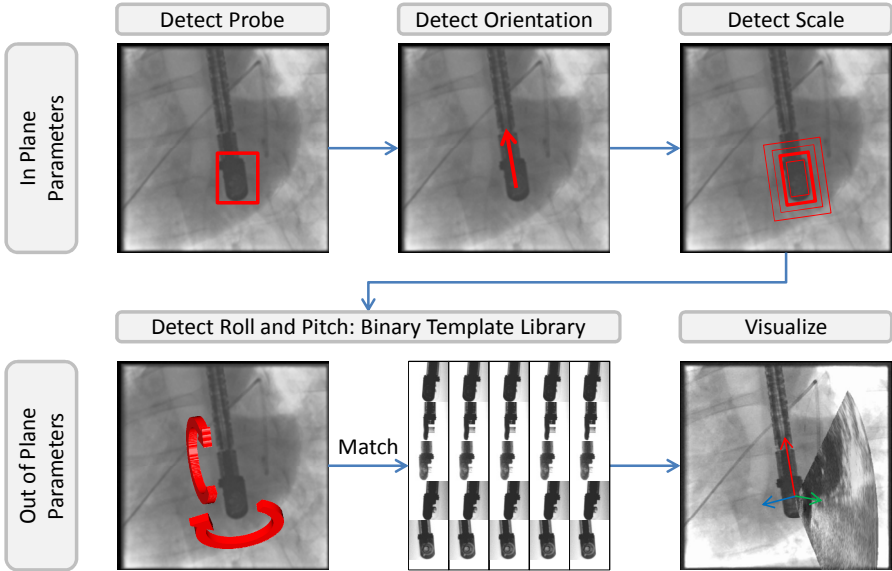
Fig. 1. Detecting the pose of a TEE probe from a single Fluoro image

2.1   TEE Probe Pose Detection

At the heart of our approach is the separation of the pose parameters into in-plane (x, y, z) and (θy) and out-of-plane $(\theta r, \theta p)$ parameters (shown in **Fig. 1**). By marginalizing the estimation problem, in-plane parameters can be efficiently estimated directly from the Fluoro images I^{Fluoro}, while being invariant to the out-of-plane parameters that are more challenging to determine.

The in-plane parameters can be computed from the probe's position (u, v), size (s) and orientation (θy) in the Fluoro image, the projection transformation $P_{projection}$ of the Fluoro device and the physical dimensions of the TEE probe. To detect the in-plane parameters (u, v), (s), (θy) from a Fluoro image I^{Fluoro} we use discriminative learning methods as described in the next section.

The out-of-plane parameters are more challenging to estimate. The visual appearance in Fluoro of the probe varies greatly making it challenging to learn a compact classifier. This requires the problem to be treated in a fundamentally different way. A template library is created of the probe's appearance under out-of-plane orientations $(\theta r, \theta p)$. Each template has an associated $(\theta r, \theta p)$ and by matching the Fluoro image to the template the out-of-plane parameters can be estimated.

Detecting In-plane Parameters

The in-plane parameters are estimated using discriminative learning methods. A classifier is trained to detect the position (u,v), size (s) and orientation (θy) of the TEE probe in the Fluoro image. The classifiers are trained using manually annotated Fluoro data. They are trained and applied sequentially such that first, candidates are detected for (u,v), then the orientation (θy) is detected for each candidate and finally the size of the probe is detected (s). Each detector is trained using a Probabilistic Boosting Tree (PBT) with Haar-like and steerable features [6].

The position (u,v) detector is trained on manual annotations and negative examples taken randomly from the Fluoro image. The Fluoro image is resized to 128×128 and a window of 35×35 is centered at the annotation. A pool of 100,000 Haar features are used to train the PBT. The appearance of the probe varies greatly and to avoid over fitting a classifier is created which is less discriminative but highly probably to detect the tip of the probe.

The orientation (θy) detector is trained on manually annotated data and the false positives from the position detector. Additional negative training data is created, centered on the annotation but with incorrect orientation parameters. The PBT is trained with five features including the relative intensity and the difference between two steerable filters [6]. The orientation detector is trained at intervals of 6° with 360° coverage. This detector is more discriminative than the position detector and therefore removes outliers as well as estimating the orientation.

The size (s) detector is trained to detect two points where the tip of the probe meets the shaft. The PBT is trained using Haar features. During detection the orientation and position of the probe are used to constrain the search area for the size detector.

Detecting Out-of-plane Parameters

The appearance of the probe under roll and pitch $(\theta r,\theta p)$ varies significantly in the Fluoro image and cannot generally be accounted for in the image space using the same techniques as the in-plane parameters. The out-of-plane parameters must be treated in a fundamentally different way. The proposed solution is to build a template library containing Fluoro images of the probe under different $(\theta r,\theta p)$. The $(\theta r,\theta p)$ parameters are estimated by matching an image patch in I^{Fluoro} (normalized for the in-plane parameters) with the template library.

A comprehensive template library should contain a wide variety of orientations. It is not feasible to build this library from *in vivo* data as it is challenging to manually annotate $(\theta r,\theta p)$ and the data may not contain complete coverage of the parameter space. The library is constructed using DRR. DRR's simulate x-ray Fluoro by tracing light rays through a 3D volume. In this work a DynaCT of the TEE probe is acquired ($512\times512\times488$ 0.2225 mm resolution). The orientation and position of the probe was manually annotated and $(\theta r,\theta p)$ rotations are applied to the volume.

Searching a large template library can be computationally expensive. The size of the library is limited to reduce the search space. The probe is not free to move in all directions due to the physical constraints of the tissue. In addition the X-ray image, formulated by integrating light, makes objects appear the same under symmetrical poses. This is exploited to reduce the size of the template library. The library is built

with pitch (θp) from -45° to 45° and roll (θr) from -90° to 90° at 2° intervals. This subsample library is still large and expensive to store and search. To make the problem computationally tractable a binary template representation is used [7],[8]. Binary templates are an efficient way of storing discriminative information for fast matching.

The image patch is divided into sub-regions and features are extracted for each region. The dominant orientation [7] of the gradient in the sub-region is used as a feature. This has been shown to work well on homogenous regions and objects which lack texture as is the case for the TEE probe in the Fluoro image. The orientations are discretized into 8 orientation bins. Each sub-region can be represented as a single byte which corresponds to the 8 orientation bins. The bit is set to 1 if the orientation exists in the sub-region and 0 if it does not. The binary template for the image patch is comprised of a set of bytes corresponding to the sub-regions. The resulting template is a compact and discriminative representation of the image patch.

Input templates extracted from the Fluoro image $F(I^{Fluoro})$ are matched to templates in the library $F(O)$ using

$$\varepsilon_1(I^{Fluoro}, O, c) = \sum_r \delta\Big(F(I^{Fluoro}, (u,v) + r) = F(O, r)\Big)$$

where $\delta()$ is a binary function which returns true if the features in two regions match, $F(I^{Fluoro}, (u,v) + r)$ is the input template centered on candidate (u,v) in image I^{Fluoro}, $F(O,r)$ is a template in the library and r is the sub-region. The function counts how many sub-regions in two templates are the same. The template in the library with the highest count is taken to be the best match and the associated $(\theta r, \theta p)$ as the out-of-plane parameters. This function can be evaluated very quickly using a bitwise AND operation followed by a bit count enabling the library to be searched efficiently.

3   Results

The proposed method for probe pose detection was validated on synthetic, phantom and *in vivo* datasets. Throughout our experiments a GE TEE Transducer was used. The synthetic dataset includes 4050 simulated Fluoro images (DRR) from a 3D C-arm Volume (DynaCT - $512\times512\times488$ 0.2225 mm pixel spacing) of the TEE probe. The ground-truth was generated by annotating the 3D probe position in the DynaCT volume. The phantom dataset includes a volumetric DynaCT of the TEE probe inserted into a silicon phantom, and a total of 51 Fluoro (960×960 0.184 mm pixel spacing) images captured by rotating the C-arm and with the TEE probe remaining static.

The position of the C-arm is known from the robotic control, which enabled ground-truth to be computed for each Fluoro image using the 3D probe annotation. The *in vivo* dataset was acquired during several porcine studies and includes 50 Fluoro sequences comprising of around 7,000 frames (512×512 0.345 mm pixel spacing). The data contains images with background clutter, catheter tools and variety in the pose of the probe, C-arm angulations, dose and anatomy. The pose parameters were manually annotated in all frames and assumed as ground-truth for training and testing.
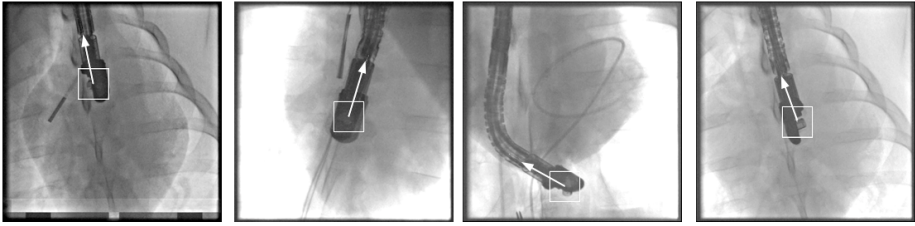
Fig. 2. Fluoroscopic images illustrating probe detection and estimation of in-plane parameters

In the first experiment the quantitative and qualitative performance evaluation of the in-plane parameter (u, v, θ_y) detection was performed on all three datasets. The detector was trained on 75% of the *in vivo* dataset (36 sequences of 5,363 frames) and tested on the entire synthetic, phantom and remaining 25% of the *in vivo* dataset. The results are summarized in Table 1.

For the *in vivo* data the average in-plane position (u, v) error was 2.2 and 3.7 mm and the in-plane orientation error was 6.69°. Errors in the position estimation are caused by false detections along the shaft of the probe. False position detections contribute to errors in the orientation estimation. The true positive rate is 0.88 and the false positive rate is 0.22. The detection and accuracy is affected by dose level, proximity to dense tissue and background clutter. The detection framework performs best when the probe is clearly distinguishable from its background. Fig. 2 illustrates detection examples and nature of *in vivo* images with cluttered background and low textured probe.

The results for the phantom and synthetic data are provided in Table 1 where detection was performed at a fixed scale. The Fluoro data from the phantom experiment appears different from the *in vivo* data used to train the detectors making it challenging. The true positive rate was 0.95 and false positive rate 0.05. False detections were caused by the density of the silicon phantom, which obscures the probe in three images. The true positive and false positive rates for synthetic data were 0.99 and 0.01 respectively. The visual appearance of the synthetic DRR is similar to the training data and the probe is clearly distinguishable causing high true positive rate.

Table 1. Quantitative validation of the in-plane position (u, v) and orientation (θy).

Data	Average Error		
	u (mm)	v(mm)	(θy)
Synthetic	1.1 (1.1)	2.2 (3.9)	2.6 (3.2)
Phantom	1.6 (1.4)	2.0 (1.2)	3.0 (3.4)
In Vivo	2.2 (5.1)	3.7 (8.0)	6.6 (16.7)
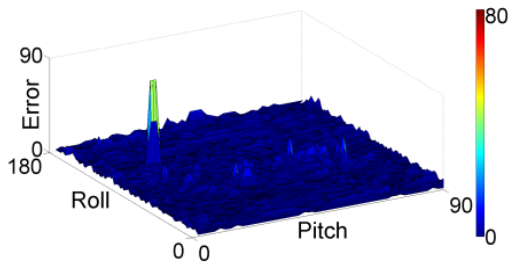
Fig. 3. Error analysis (degrees) of $(\theta r, \theta p)$ over search space
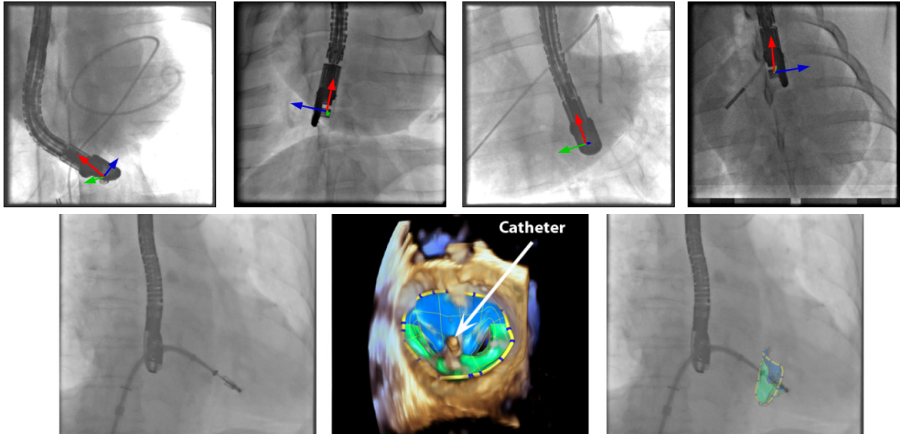
Fig. 4. Top - Fluoro images showing the detected pose of probe. Bottom: Left- Fluoro image. Center – mitral valve detected in 3D TEE. Right – Valve model visualized in Fluoro.

The out-of-plane $(\theta r, \theta p)$ detectors are analyzed on the synthetic data to evaluate the accuracy of the binary template matching. Fig. 3 plots the $(\theta r, \theta p)$ error over the search space (degrees) and illustrates stable detection with a single outlier.

The framework is evaluated with respect to all parameters (Table 2). Quantitative validation was performed on synthetic and phantom data (*in vivo* ground truth data was not available). The largest error is in the Z axis, which corresponds to the optical axis of the Fluoro device. It is expected that this is the largest error because estimating distance along the optical axis is challenging from a monocular Fluoro image. Fortunately, the goal of the framework is to visualize anatomy in the Fluoro image, therefore errors in Z has little effect on the final visualization. Initial clinical feedback suggests errors of up to 15° and 10 mm (excluding Z) are acceptable for some visualizations, however accuracy requirements are application specific. Qualitative evaluation (Fig. 4 top) is performed on *in vivo* Fluoro images.

Table 2. Quantitative validation of TEE probe detection

Data	X (mm)	Y(mm)	Error Z(mm)	(θr)	(θp)	(θy)
Synthetic	0.82 (0.79)	0.97 (2.1)	64.0(13.9)	4.2 (10.5)	4.6 (9.0)	2.6(3.2)
Phantom	1.1 (0.8)	0.7(0.6)	19.04(1.6)	11.5(12.0)	11.8(9.8)	3.0(3.4)

The computational performance was evaluated (Intel 2.13GHz single core, 3.4GB RAM). The average detection time is 0.53 seconds. The computational cost can be reduced by incorporating temporal information to reduce the search space.

To illustrate the clinical relevance of this work an anatomical model of the mitral valve is detected [9] in 3D TEE and visualized in Fluoro (**Fig. 4** bottom). The data is not synchronized and is manually fused. A catheter is visible in both modalities.

4   Conclusions

This paper presents a novel method for automated fusion of TEE and Fluoro images to provide guidance for cardiac interventions. The proposed system detects the pose of a TEE probe in a Fluoro image. Discriminative learning is combined with fast binary template matching to address the challenges of pose detection. Validation has been performed on synthetic, phantom and *in vivo* data. The method is capable of detecting in 0.5s with an in-plane accuracy of less than 5 mm. Future work will focus on incorporating temporal information, using the initial detected pose as a starting estimate for pose refinement and visualization of anatomically meaningful information.

References

1. Jain, A., Gutierrez, L., Stanton, D.: 3D TEE Registration with X-Ray Fluoroscopy for Interventional Cardiac Applications. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 321–329. Springer, Heidelberg (2009)
2. Ma, Y., Penney, G.P., Bos, D., Frissen, P., Rinaldi, C.A., Razavi, R., Rhode, K.S.: Hybrid echo and x-ray image guidance for cardiac catheterization procedures by using a robotic arm: a feasibility study. Physics in Medicine and Biology 55, 371–382 (2010)
3. Gao, G., Penney, G., Ma, Y., Gogin, N., Cathier, P., Arujuna, A., Morton, G., Caulfield, D., Gill, J., Aldo Rinaldi, C., Hancock, J., Redwood, S., Thomas, M., Razavi, R., Gijsbers, G., Rhode, K.: Registration of 3D trans-esophageal echocardiography to X-ray fluoroscopy using image-based probe tracking. Medical Image Analysis 16, 38–49 (2012)
4. Gao, G., Penney, G., Gogin, N., Cathier, P., Arujuna, A., Wright, M., Caulfield, D., Rinaldi, A., Razavi, R., Rhode, K.: Rapid Image Registration of Three-Dimensional Transesophageal Echocardiography and X-ray Fluoroscopy for the Guidance of Cardiac Interventions. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 124–134. Springer, Heidelberg (2010)
5. Lang, P., Seslija, P., Chu, M.W.A., Bainbridge, D., Guiraudon, G.M., Jones, D.L., Peters, T.M.: US-Fluoroscopy Registration for Transcatheter Aortic Valve Implantation. IEEE Transactions on Biomedical Engineering 59, 1444–1453 (2012)
6. Wu, W., Chen, T., Wang, P., Zhou, S.K., Comaniciu, D., Barbu, A., Strobel, N.: Learning-based hypothesis fusion for robust catheter tracking in 2D X-ray fluoroscopy. In: CVPR, pp. 1097–1104 (2011)
7. Hinterstoisser, S., Lepetit, V., Ilic, S., Fua, P., Navab, N.: Dominant orientation templates for real-time detection of texture-less objects. In: CVPR, pp. 2257–2264 (2010)
8. Taylor, S., Drummond, T.: Multiple Target Localisation at over 100 FPS. In: BMVC (2009)
9. Ionasec, R.I., Voigt, I., Georgescu, B., Wang, Y., Houle, H., Vega-Higuera, F., Navab, N., Comaniciu, D.: Patient-specific modeling and quantification of the aortic and mitral valves from 4-D cardiac CT and TEE. IEEE Trans. on Med. Imag. 29, 1636–1651 (2010)

Direct 3D Ultrasound to Video Registration Using Photoacoustic Effect

Alexis Cheng, Jin U. Kang, Russell H. Taylor, and Emad M. Boctor

The Johns Hopkins University, Baltimore, Maryland 21211, United States of America
{acheng22,jkang,rht}@jhu.edu, eboctor1@jhmi.edu

Abstract. Interventional guidance systems require surgical navigation systems to register different tools and devices together. Standard navigation systems have various drawbacks leading to target registration errors (TRE) of around 3mm. The aim of this work is to introduce the photoacoustic (PA) effect as a direct 3D ultrasound (US) to video registration method. We present our experimental setup and demonstrate its feasibility on both a synthetic phantom and an *ex vivo* tissue phantom. We achieve an average TRE of 560 microns and standard deviation of 280 microns on a synthetic phantom. Also, an average TRE of 420 microns and standard deviation of 150 microns on the *ex vivo* tissue phantom are obtained. We describe a roadmap to bring this system into the surgical setting and highlight possible challenges along the way.

Keywords: Instrument & Patient Localization and Tracking, Registration, Planning and image guidance of interventions, Visualization, Abdominal imaging, Oncology Applications, Optical imaging, Ultrasound.

1   Introduction

Interventional guidance systems are commonly used in modern surgical procedures, including open surgery, laparoscopic surgery, and robotic surgery [1]. The guidance systems are used because it is often difficult to locate tumors both due to its movement during the procedure and the camera's limited field of view. Guidance systems typically provide a fusion of video and other imaging modalities to reduce the difficulty in locating regions of interest. Before they can be used, these guidance systems must be registered with a number of surgical tools and devices. To accomplish this registration, several guidance systems are equipped with electromagnetic (EM) or optical surgical navigation systems [2], [3].

Stolka et al. [3] present an EM-based laparoscopic intra-operative ultrasound navigation system similar to standard EM-based systems shown in Figure 1A. These systems typically have several drawbacks including the number of EM sensors that must be placed on all tools in order to be tracked by the navigation system and the need to indirectly transform from one coordinate frame to another. An indirect transformation is defined to be one where the transformation to the coordinate frame of interest cannot be obtained without computing a chain of transformations via intermediate coordinate

frames. As an example, to obtain the transformation from EM space to the US image space, the transformation from the EM frame to the marker on the US probe and the transformation from the marker to the US frame are needed. Determining this second transformation presents another problem of calibration. A number of authors [4], [5] have presented research on specific calibration processes, attempting to minimize the error. Typically, they have demonstrated errors of around 3mm [4], [6], [7].

Yip et al. [8] have demonstrated a registration method that utilizes a tool serving as both fiducials in the US space and markers in the stereo camera (SC) space. Some drawbacks for this system include the need to have a custom registration tool touching the surface and the segmentation of fiducials from noisy US images.

Vyas et al. [9] demonstrated proof of concept for a direct registration method with the PA effect. For this paper, a direct registration method is defined to be one that only requires a single transformation between frames as opposed to a chain of transformations. The method addresses each of the drawbacks above. A marker is not necessary to generate a coordinate transformation between the tracker frame and the US transducer frame. Previous work [10], [11] shows that a pulsed laser source can effectively generate the photoacoustic (PA) signal in tissue, resulting in an acoustic wave that can be detected by conventional US transducers [12], [13]. Other than the US transducer, nothing else needs to touch the surface. Each laser point projection can be seen as a green spot in the SC space and as a PA signal in the US space. Segmentation of the PA signal is also simpler in a PA image because the laser spot is the only acoustic source. Finally, the calibration process is unnecessary since the coordinate transformation from the SC frame to the US frame can be computed directly with the two 3D point sets based on rigid registration algorithms [2], [16].
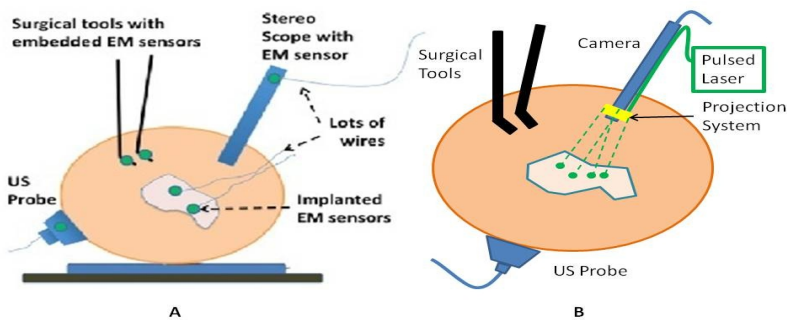
Fig. 1. A) Standard EM-based Navigation System, B) PA Navigation System

Our work introduces the PA effect and PA imaging as a new medical tracking technology and is a step towards realizing the system shown in Figure 1B. We aim to present a direct 3D US to video registration method and to demonstrate its feasibility on *ex vivo* tissue. Improving on [9], we use a 3D US transducer instead of a 2D US transducer to detect the PA signal. Using a 3D transducer allows this registration method to function for a non-planar set of 3D points. This is a significant advantage because we aim to deploy this method in a laparoscopic environment and organ surfaces will rarely form a planar surface. We also improve significantly on the point

finding algorithms used in [9] for both SC images and US volumes. In addition to using a synthetic phantom with excellent light absorption characteristics, we also use a piece of resected *ex vivo* porcine liver tissue embedded in a gelatin phantom to demonstrate this method's feasibility in a practical environment and eventual deployment in our intended applications of laparoscopic tumor resections.

This paper will detail the experimental procedure and algorithms to validate this method on a synthetic phantom and an *ex vivo* liver phantom using a 3D US transducer. We will present target registration error (TRE) results. Possible challenges in bringing this system into the surgical setting will also be discussed.

2   Methods

To perform our experiment, we use a Q-switched neodymium-doped yttrium aluminum garnet (Nd:YAG), Brilliant (Quantel Laser, France) laser frequency doubled to 532nm wavelength at approximately 6mJ/cm^2 to generate a PA effect on the synthetic phantom and approximately 19mJ/cm^2 on the *ex vivo* tissue phantom. At this wavelength, most of the laser energy is absorbed at the superficial surface of the tissue. However, there is slight penetration into the tissue, creating a source of error that will be discussed. Our stated energy is lower than the maximum permissible exposure of 19.5mJ/cm^2 as calculated from the IEC 60825-1 laser safety standard [14] based on a 0.25s exposure time, a 4ns pulse width, and a frequency of 10Hz. Alternate tests showed that a lower energy was also able to generate a PA effect on *ex vivo* tissue. We use a SonixCEP US system along with a 4DL14-5/38 US transducer developed by Ultrasonix Medical Corporation (Richmond, Canada) to scan the volume of interest. The motor actuation of this transducer induces angular movement around an internal pivot point. The Sonix DAQ device, developed in collaboration between the University of Hong Kong and Ultrasonix, and the MUSiiC toolkit [15] is used to acquire pre-beamformed radiofrequency (RF) data directly from the US machine. We use the k-wave toolbox [16] in MATLAB (Mathworks Inc. Natick, MA) designed for reconstructing PA images based on RF data. A custom-built SC system containing two CMLN-13S2C cameras (Point Grey Research, Richmond, Canada) is used to capture images to be used for 3D triangulation. The synthetic phantom is made using plastisol and black dye. The *ex vivo* liver phantom is made using a gelatin solution and a freshly resected porcine liver. The surface of the liver is partially exposed and not covered by gelatin. Alternate tests with other surfaces such as porcine kidney tissue and fat were also successful in generating a PA signal.

Our experiment can be split into a data collection phase and a data processing phase. The data collection phase outputs SC image pairs, five frames for each camera, and a 3D RF US volume for each projected laser spot. The number of frames is arbitrary. The data processing phase uses the data and generates a coordinate transformation from the SC frame to the US frame. Figure 2A shows the experimental setup and an overlay of a US image representation using the inverse of the transformation.

Figure 3A shows the workflow of the data collection phase. First a laser spot is projected onto the exposed surface of the *ex vivo* liver phantom. There will be inaccuracies

in SC spot triangulation if the laser spot is projected at or near the liver gelatin interface because the laser spots become irregularly shaped when projected onto clear materials. Second, several images are taken for each camera. The laser spot projected onto the phantom must be visible in at least one image per camera for triangulation to be possible. Our cameras have a faster capture rate than our laser's repetition rate, so some of the frames will be devoid of the laser signal. We will exploit this during data processing. Steps 3 and 4 show that the 3D US transducer motor actuation and RF data are intermittently collected from the DAQ device to scan and acquire the RF data of the volume of interest. The volume's field of view is 14.7^o for the *ex vivo* tissue phantom, 19.6^o for the synthetic phantom and the motor step size is 0.49^o. This iterative process is manual, but an automatic process is in development. This workflow is repeated for each of the thirty laser spots.
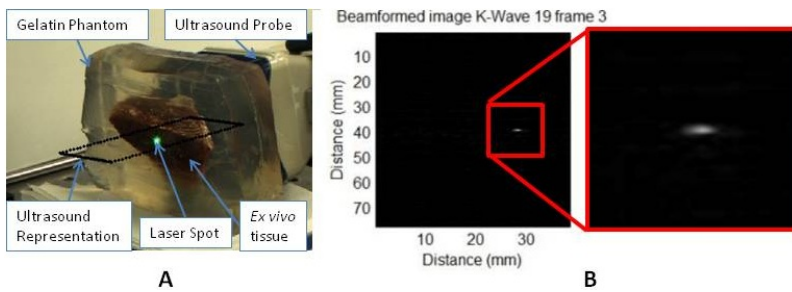
Fig. 2. A) Experimental Setup and Video Overlay, B) PA Signal within an US image
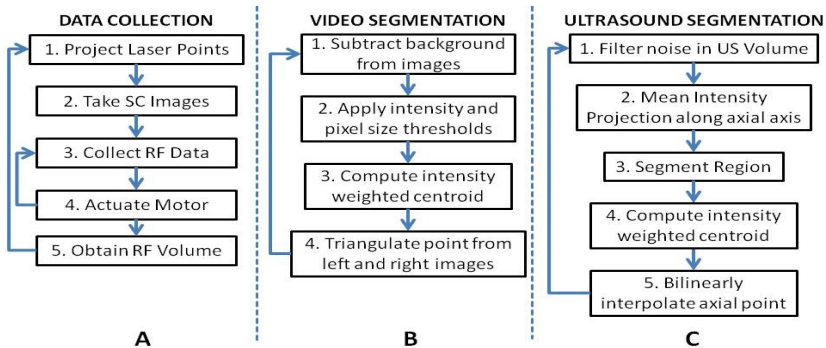
Fig. 3. Workflow for A) Data Collection, B) SC Segmentation, and C) US Segmentation

The data processing phase involves the segmentation of the SC images into 3D SC points, the segmentation of the 3D RF US volume data into 3D US points, and the computation of the transformation from the SC frame to the US frame.

Figure 3B shows the workflow for SC segmentation. For each camera, we pick a SC image with the laser spot and without the laser spot. Next, the background images without the laser spot are subtracted from the images with the laser spot. This step makes it significantly easier to segment the laser spot. We then apply an appropriate intensity and pixel size thresholds such that the laser spot is segmented out.

These thresholds are selected based on the laser beam diameter and the phantom's reflectance. Next, we fit an ellipse to the segmented region and compute the intensity weighted centroid. Calibration files for our specific SC allow us to triangulate the segmented point from each camera and obtain a single 3D point in the SC frame. This workflow is repeated for each laser spot projection. We use thirty sets of SC images.

The workflow for the segmentation of the 3D RF US volume is shown in Figure 3C. First, for each slice of a 3D RF US volume, the RF data is beamformed using the k-wave toolbox [16] in MATLAB. The dynamic range of the image is normalized with respect to the volume to decrease the size of the PA signal seen in each volume. Figure 2B shows the the k-wave beamformed PA signal image. Next, we project the volume onto the lateral-elevational plane by taking the mean along each axial ray. An appropriate intensity and pixel size threshold is then applied to this image. An ellipse is fitted on the segmented region and an intensity-weighted centroid is computed resulting in lateral and elevational coordinates. As described earlier, the PA signal originates from the surface and any penetration into the surface. Since air cannot generate a PA signal in our setup, we can exploit that the high intensity pixels farthest away in the axial direction are from the surface. Thus, we obtain the axial coordinate corresponding with a lateral-elevational coordinate as the axial-most high intensity pixel. This step is particularly important because the penetration of the laser pulse is much deeper for the *ex vivo* tissue phantom than the penetration for the synthetic phantom. We use bilinear interpolation to obtain axial coordinates between sampled points. These three coordinates are converted to 3D US coordinates based on transducer specifications. This workflow is repeated for each of our thirty 3D RF US volumes.

The transformation from the SC frame to the US frame can be computed with the 3D SC and 3D US point sets. Any registration method for computing the transformation between two 3D point sets can be used. We use the coherent point drift algorithm [17] in our experiment. One of the main reasons for using coherent point drift is that it allows for data points to be missing from either dataset. An assumption that we have made is that each laser spot will be visible in the SC images and each PA signal will be visible in the US volume. This assumption is valid for our experiment, but may not hold in the surgical setting due to SC or transducer movement. The coherent point drift registration algorithm allows us to acquire a registration as long as there are enough corresponding points in the SC images and the US volume.

The transformation from the SC frame to the US frame is used to transform the 3D SC points to the US frame for validation. The inverse transformation is used to display a representation of an US image into the SC frame as shown in Figure 2A.

3   Results

The results of our experiment on the synthetic phantom and on the *ex vivo* tissue phantom are validated using the target registration error (TRE) metric defined in equation 1. F_{SC_US} is the transformation from the SC frame to the US frame and is computed with all of SC and US points except for one. The TRE is the resulting difference between the actual US test point and the transformed SC test point in the US frame.

$$\vec{TRE} = \mathbf{F}_{SC\_US} * \vec{SC}_{test} - \vec{US}_{test} \qquad (1)$$

Twenty-nine of the thirty points are used to compute the transformation from the SC frame to the US frame. The remaining point is used as a test point to compute the TRE. This computation is repeated with each of the thirty points as test points. Table 1 shows the average and standard deviation of the TRE results for the thirty cases in the synthetic phantom and the *ex vivo* tissue phantom experiment respectively.

Table 1. Average TRE Results for Leave One Out Registration Experiments

n = 30	Synthetic Phantom	*Ex vivo* Tissue Phantom
Lateral (mm)	0.21 ± 0.17	0.22 ± 0.16
Axial (mm)	0.21 ± 0.13	0.24 ± 0.15
Elevational (mm)	0.41 ± 0.31	0.18 ± 0.10
Euclidean Norm (mm)	0.56 ± 0.28	0.42 ± 0.15

4   Discussion

There are several considerations when discussing this system's deployment in our intended applications of laparoscopic tumor resections. The first is the placement of the transducer. In our experiments, we use a relatively large 3D US transducer that would be near impossible to put inside the body during a laparoscopic procedure. However, the transducer is often placed externally [3], [8] in these procedures, so the size of the probe is not an issue. Naturally, there are disadvantages of placing the transducer externally and farther from the region or organ of interest. The quality of ultrasound images degrades as the depth increases, which would likely lead to errors in localizing fiducials or, in our case, the PA signal. However, since the PA signal only has to travel in one direction, as opposed to traditional US, our PA images will have better quality than US images of equivalent depth.

Another issue with our 3D US transducer is the acquisition speed. There are certain applications where an acquisition speed of a volume per several seconds is sufficient, but a real-time implementation would require a higher acquisition rate. We anticipate using 2D array US transducers for a real-time implementation. These transducers would provide an acquisition rate on the order of twenty volumes per second. The 2D array transducer could also be miniaturized and placed closer to the region of interest.

A third issue deals with the laser delivery system. As shown in our experimental setup, a laser would have to be fired at the organ in free space. This occurrence is unlikely in practical situations. We are developing a fiber delivery tool that will allow us to safely guide the laser beam into the patient's body. This tool will also be able to project concurrent laser spots, greatly enhancing our registration acquisition rate.

At the level of error measurements shown in table 1, it is likely that the calibration of the SC system is a significant contributor. They are able to locate point sources at sub-millimeter accuracy [6], [7]. This error is usually negligible in comparison with the 3mm errors from calibration. Since our results are 0.56mm and 0.42mm errors

respectively, the SC system's error becomes significant. We use a custom SC system, so its errors are likely greater than a finely tuned commercial SC system.

The experimental results in table 1 show that our system achieves sub-millimeter TRE measurements for both the synthetic phantom and the *ex vivo* tissue phantom. There is a slight difference in the results, and it is entirely due to the elevational error. This is likely due to the larger field of view in the synthetic phantom experiment as well as normal variation across experiments. More experiments must be performed to obtain an average error across multiple experiments.

There are a couple of factors that will affect these errors as we move from a bench-top setup to *in vivo* experiments. When our SC system is replaced with a stereo endoscopic camera, the errors may increase. This is because our SC system has a larger disparity than standard stereo endoscopic cameras. Further work will be done to quantify the effects of this change. Also, the errors are reported based on surface points. Since the region of interest is often subsurface, our reported TRE will be biased for subsurface target errors. We believe that the bias will be fairly small since the PA spots are being detected in the same modality as any subsurface regions.

5   Conclusion

We have proposed an innovative 3D US to video direct registration medical tracking technology based on the PA effect and demonstrated its feasibility on an *ex vivo* tissue phantom. The results have been shown to have higher accuracy than state of the art surgical navigation systems. Future work will include the development of a fiber delivery tool, spot finding algorithms to support concurrent spot projection, and subsequent *in vivo* animal experiments. Integration of this direct registration method into laparoscopic or robotic surgery environments will also be a point of emphasis.

References

1. Wang, Y., Butner, S., Darzi, A.: The developing market for medical robotics. Proceedings of the IEEE 94(9), 1763–1771 (2006)
2. Taylor, R., Lavallee, S., Burdea, G., Mosges, R.: Computer integrated surgery. MIT Press, Cambridge (1996)
3. Stolka, P.J., Keil, M., Sakas, G., McVeigh, E., Allaf, M.E., Taylor, R.H., Boctor, E.M.: A 3D-elastography-guided system for laparoscopic partial nephrectomies. In: Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling, San Diego, February 13-18, vol. 76251I, p. 76251I-12 (2010)

4. Boctor, E., Viswanathan, A., Choti, M., Taylor, R., Fichtinger, G., Hager, G.: A Novel Closed Form Solution for Ultrasound Calibration. In: International Symposium on Biomedical Imaging, Arlington, pp. 527–530 (2004)
5. Poon, T., Rohling, R.: Comparison of calibration methods for spatial tracking of a 3-D ultrasound probe. Ultrasound in Medicine and Biology 31(8), 1095–1108 (2005)
6. Navab, N., Mitschke, M., Schuetz, O.: Camera-Augmented Mobile C-arm (CAMC) Application: 3D Reconstruction Using a Low-Cost Mobile C-arm. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 688–697. Springer, Heidelberg (1999)
7. Wiles, A., Thompson, D., Frantz, D.: Accuracy assessment and interpretation for optical tracking systems. In: Proceedings of SPIE, vol. 5367, pp. 421–432 (2004)
8. Yip, M.C., Adebar, T.K., Rohling, R., Salcudean, S.E., Nguan, C.Y.: 3D Ultrasound to Stereoscopic Camera Registration through an Air-Tissue Boundary. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 626–634. Springer, Heidelberg (2010)
9. Vyas, S., Su, S., Kim, R., Kuo, N., Taylor, R.H., Kang, J.U., Boctor, E.M.: Interoperative Ultrasound to Stereocamera Registration using Interventional Photoacoustic Imaging. In: Medical Imaging 2012: Visualization, Image-Guided Procedures, and Modeling, San Diego, February 4-9, vol. 8316, p. 83160S (2012)
10. Kolkman, R., Steenbergen, W., van Leeuwen, T.: In vivo photoacoustic imaging of blood vessels with a pulsed laser diode. Lasers in Medical Science 21(3), 134–139 (2006)
11. Kuo, N., Kang, H.J., DeJournett, T., Spicer, J., Boctor, E.M.: Photoacoustic imaging of prostate brachytherapy seeds in ex vivo prostate. In: Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling, Lake Buena Vista, February 12-17, vol. 7964, p. 796409 (2011)
12. Xu, M., Wang, L.: Photoacoustic imaging in biomedicine. Review of Scientific Instruments 77, 041101 (2006)
13. Hoelen, C., De Mul, F., Pongers, R., Dekker, A.: Three-dimensional photoacoustic imaging of blood vessels in tissue. Optics Letters 23(8), 648–650 (1998)
14. IEC 60825-1:1993+A1:1997+A2:2001: Safety of Laser Products – Part 1: Equipment Classification and Requirements. International Electrotechnical Commission, Geneva (2001)
15. Kang, H.J., Kuo, N., Guo, X., Song, D., Kang, J.U., Boctor, E.M.: Software framework of a real-time pre-beamformed RF data acquisition of an ultrasound research scanner. In: Medical Imaging 2012: Visualization, Image-Guided Procedures, and Modeling, San Diego, February 4-9, vol. 8320, p. 83201F (2012)
16. Treeby, B., Cox, B.: k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields. Journal of Biomedical Optics 15(2), 21314 (2010)
17. Myronenko, A., Song, X.: Point-Set Registration: Coherent Point Drift. IEEE Trans. on Pattern Analysis and Machine Intelligence 32(12), 2262–2275 (2010)

Assessment of Navigation Cues with Proximal Force Sensing during Endovascular Catheterization

Hedyeh Rafii-Tari[1], Christopher J. Payne[1], Celia Riga[2], Colin Bicknell[2], Su-Lin Lee[1], and Guang-Zhong Yang[1]

[1] The Hamlyn Centre for Robotic Surgery
[2] Academic Division of Surgery
Imperial College London, UK
{h.rafii-tari11,christopher.payne,g.z.yang}@imperial.ac.uk

Abstract. Despite increased use of robotic catheter navigation systems for endovascular intervention procedures, current master-slave platforms have not yet taken into account dexterous manipulation skill used in traditional catheterization procedures. Information on tool forces applied by operators is often limited. A novel force/torque sensor is developed in this paper to obtain behavioural data across different experience levels and identify underlying factors that affect overall operator performance. The miniature device can be attached to any part of the proximal end of the catheter, together with a position sensor attached to the catheter tip, for relating tool forces to catheter dynamics and overall performance. The results show clear differences in manipulation skills between experience groups, thus providing insights into different patterns and range of forces applied during routine endovascular procedures. They also provide important design specifications for ergonomically optimized catheter manipulation platforms with added haptic feedback while maintaining natural skills of the operators.

Keywords: endovascular intervention, catheter manipulation, skill assessment, robotic catheterization, force sensing.

1 Introduction

Robotically controlled steerable catheter navigation systems have seen a growing interest in the field of endovascular surgery. These systems offer potential advantages over manual catheterization, including reduced radiation exposure, increased precision and stability of motion, and added operator comfort [1]. However, most of these systems have been designed with little consideration of underlying perceptual cues, operator-tool force interaction and behavioural patterns, thus not fully utilizing natural ergonomic skills used during conventional catheterization that are obtained through experience. A clear understanding of these force and motion patterns and manipulation skills is crucial for designing next generation intuitive catheter navigation systems by maximizing natural bedside catheterization skills of operators.

In practice, manual catheterization is based on 2D visual guidance (fluoroscopy) and haptic cues to sense small axial forces and torques at the fingertips while manipulating catheters and guidewires. This is achieved using a combination of pushing, pulling, and twisting at the proximal end of the tools in different directions based on an implicit model of the catheter acquired through experience and a mental picture of the 3D anatomy augmented with 2D real-time image data. Understanding the forces and torques that are applied during a procedure is important to avoid injuries that can be caused by the interactions of catheters and guidewires with the vessel walls, especially in high risk areas and lesions that may lead to perforation or thrombosis.

Thus far, one of the main commercial interventional robots is the Sensei robotic navigation system (Hansen Medical, Mountain View, CA, USA), which has been successfully used in different clinical applications including cardiac ablation and endovascular aneurysm repair [2, 3]. While its efficacy in reducing radiation exposure and fluoroscopy time has been validated [2], its limitations include its large size, high cost, and longer setup times. In the research domain, different groups have also focused on telerobotic master and slave systems for catheter navigation, with added proximal or distal force sensing and haptic force feedback to the operator [4–7]. For most of these systems, the master takes the shape of a joystick or a haptic device (e.g. Novint's Falcon, Phantom Omni), therefore forgoing the natural catheter/guidewire manipulation skills and haptic cues used during bedside practice. Studies show that clinical success is highly dependent on operator experience, and endovascular procedures are associated with steep learning curves [8]. This highlights the importance of maintaining operator skills and dexterity for successful robotic catheter navigation. It also motivates the development of remote catheter navigation systems that maintain the conventional manipulation skills through the replication of motion, force and sensation during the manipulation of a local catheter.

To date, few studies have looked at operator behavioural data and catheter dynamics by mostly focusing on finger motion patterns, or measuring catheter kinematics and forces to overcome introducer sheath friction [7, 9]. Direct measurement of proximal tool forces applied to the catheter and relating these to catheter tip motion can provide critical information on endovascular manipulation skills and present useful design characteristics for improved catheter navigation systems.

The purpose of this paper is to propose an endovascular navigation platform for assessing detailed navigation cues of different operators. A novel force and torque sensor attached to the proximal end of the catheter is developed together with a position sensor attached to the tip, in order to directly relate tool forces applied by operators to catheter tip motion and path length. Performance results, including subject-specific manipulation strategies, are compared over different experience levels in a realistic endovascular setting so as to gain an understanding of force and torque patterns, haptic cues, and underlying skills that contribute to overall operator performance. The study provides important design specifications for the future development of ergonomically optimized catheter manipulation

platforms with added haptic feedback, whilst taking full advantage of natural skills of the operators for endovascular intervention.

2   Materials and Methods

2.1   Force Sensor Design

The proposed force-torque (F/T) sensor measures the axial (push and pull) and torsional (clockwise and counterclockwise) loads exerted on to the catheter by the operator (Fig. 1). The sensor is miniaturized and designed to be as unobtrusive as possible to minimize the effects on catheter dynamics and operator skills; this was confirmed by the experienced operators. Instead of manipulating the catheter directly, the operator manipulates a co-axial over-tube and the force measurements are made between the over-tube and the catheter. The over-tube transmits axial and torsional loads on to a transmission component, which is seated in low friction polymer bearings within the sensor casing. The load is then transmitted on to four force sensors within the casing. A spring-loaded clamp affixes the catheter to the casing which allows the F/T measurement to be made between the over-tube and catheter. The clamp is designed so as to avoid catheter bending which would increase the friction between over-tube and catheter. By depressing the clamp, the sensor can be positioned anywhere along the length of the catheter that is comfortable to the operator.

The over-tube was chosen to have a bending, torsional stiffness, and outer diameter similar to that of the catheter so that the operator would feel as if they were manipulating the catheter directly. Additionally, to avoid errors in the measurements, the over-tube was chosen to have low friction and to have a high crush resistance, so that the user could not squeeze the catheter through the over-tube. The force sensors (FSS1500NS, Honeywell) were chosen for their compact size, low weight and linearity and were calibrated against a Nano 17 F/T sensor (ATI Industrial Automation Inc., USA).
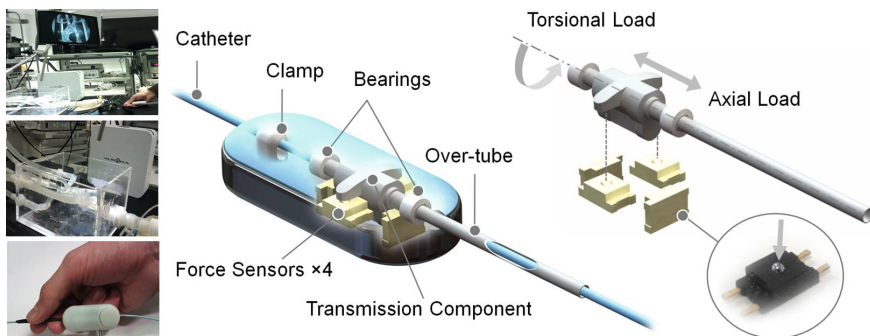
Fig. 1. Experimental setup (left) and force and torque sensor mounted on catheter with exploded view of transmission component showing the four force sensors (right)

2.2   Experimental Setup

A phantom study was performed to obtain tool force and torques applied by operators with varying endovascular skills, and to relate catheter tip motions and velocities to forces applied at the proximal end. It was also designed to extract skill related patterns and underlying factors that affect catheter navigation and overall performance within different steps of an endovascular procedure.

To obtain information on catheter tip position, velocity, and path length, a six degree of freedom electromagnetic position sensor (Aurora, ND technologies) was attached to the tip of the catheter. A silicone-based, transparent, anthropomorphic phantom (Elastrat Sarl, Geneva, Switzerland) representing a type I aortic arch was used for this study (Fig. 1). Eight subjects of varying endovascular experience were recruited and separated into two groups: experienced (n=3, more than 100 endovascular procedures) and inexperienced (n=5, 0 endovascular procedures). All operators were right-handed. Each operator was asked to cannulate the right common carotid artery of the phantom three times. Each trial was considered an independent test, thereby providing sufficient experiments for comparing the two distinct skill sets. In order to simulate 2D fluoroscopy guidance in the OR, a laparoscopic camera was mounted above the model, with the 2D image projected on a monitor to be used by operators for navigation. All operators underwent a short training session to familiarize themselves with the use of the force sensor before commencing the study. Appropriate endovascular tools, including wires and 5F shaped catheters, were available. Force measurements were read into Labview using an acquisition card (NI-USB6009, National Instruments Corp., USA) at 25 Hz.

Depending on the location of the catheter within the vasculature, force values and tip movements can vary significantly, therefore the procedure path was divided into three sections as shown in Fig. 2a : traversing the descending aorta (section A), passing through the aortic arch (section B), and finally cannulating the right common carotid artery (RCCA)(section C). Results were compared over each phase of the procedure for different experience levels.

Ten performance metrics were measured for each section of each procedure. These included median and maximum tip velocity, mean proximal forces in each axial direction, mean torques in each rotational direction, mean tip distance from the origin (catheter path length), sum of catheter tip displacements (motion efficiency), number of twists applied at the proximal end, and procedure time. To identify underlying factors that explain causalities and patterns of correlation between these observed variables, factor analysis was performed on six of these variables for each experience group. The factor loadings were extracted with a maximum likelihood estimate (MLE) for two common factors using a Varimax factor rotation. Differences between experienced and inexperienced operators were also assessed with a non-parametric Mann-Whitney U significance test on all the metrics over each section of the procedure ($P < 0.05$). Skill-related patterns of behaviour can be extracted by comparing graphs of proximal forces, torques, and catheter tip displacement between operators. Dynamic time warping was used to analyze the similarities of these parameters between an

experienced and inexperienced operator over the three different sections of the procedure. The range of forces and torques applied over all procedures are also reported for the two experience groups. Statistical analysis was performed with the SPSS software (SPSS Inc., Chicago, Il) and Matlab's statistical toolbox.

3   Results

Fig. 2 depicts the path of the catheter within a mesh of the vascular model, guided by an experienced (b) vs. inexperienced (c) operator. The color gradient depicts the magnitude of the torque applied by the operator at the proximal end of the catheter, over the catheter path.
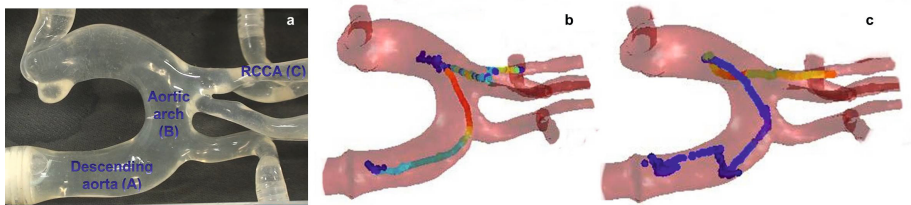
Fig. 2. Vascular phantom depicting the three sections of the procedure (a), catheter path inside the model for experienced (b) and inexperienced (c) operator, with color gradient depicting the magnitude of torque measured at the proximal end

Table 1 shows the result of the factor analysis with two extracted factors, at each section of the procedure. Each value represents the correlation between the variable and the underlying factor (the largest loadings are highlighted). For the descending aorta, mean catheter tip displacement, number of twists, and time are highly correlated for experienced operators as opposed to inexperienced operators, therefore the underlying factor could be a measure of operator efficiency and gain in motion to advance the catheter inside the aorta. In the aortic arch section, there is a high loading on torque, twisting and catheter motion for experienced operators, as compared to push force for inexperienced operators. This emphasizes trained catheter manipulation skills of operators in high-risk areas to avoid damage to the vessels. Results for the common carotid artery also depict the reliance of experienced operators on torque (rather than force) for tip displacement, in order to avoid contact with the narrow walls of the artery.

The results of the non-parametric test between experienced and inexperienced groups depict significant differences for mean displacement ($P = 0.02$), number of twists ($P = 0.02$) and time ($P = 0.05$) in the first part of the procedure (descending aorta). In the arch section, average torque was the more significant variable ($P = 0.01$), while for the RCCA section median speed ($P = 0.04$) and mean displacement ($P = 0.03$) showed significant differences between operators.

Table 2 shows the difference in performances for these significant metrics in each section of the procedure. For the descending aorta, the median values for

Table 1. Factor analysis with 2 common factors over each section of the procedure for experienced and inexperienced operators. Highly correlated factors are highlighted.

	Descending aorta				Aortic arch				RCCA			
	Exp.		Inexp.		Exp.		Inexp.		Exp.		Inexp.	
	1	2	1	2	1	2	1	2	1	2	1	2
Median speed	1.00	0.01	0.53	0.84	-0.14	-0.98	-0.59	-0.06	-0.76	0.27	-0.65	-0.60
Mean displacement	0.23	0.85	-0.26	-0.01	-0.90	0.33	0.58	0.0	0.12	0.64	0.64	-0.05
Mean push force	0.55	-0.46	0.21	-0.55	0.01	-0.31	0.37	0.47	0.33	0.19	-0.83	-0.27
Mean torque CCW	-0.87	0.05	0.17	0.20	0.44	0.82	0.04	0.35	0.13	-0.99	0.92	0.38
Number of twists	-0.19	0.63	0.75	0.13	0.68	0.33	0.05	0.93	1	0.04	0.51	0.78
Time	-0.18	0.91	0.92	0.07	0.86	0.50	0.86	0.40	0.96	0.12	-0.03	1.00

catheter displacement and number of twists show that experienced operators are more ergonomic by using less repetitive movements while achieving a higher gain in tip displacement. The difference in torque in the second part of the procedure proves our previous results that experienced operators rely much more on manipulation skills and torque when maneuvering the catheter through high-risk areas such as the arch. The difference in speed for the last section of the procedure shows cautious and smooth navigation skills of experienced operators to achieve slower yet more efficient catheter motion through narrow arteries.

Table 2. Median values of statistically significant parameters for the two experience groups at each section of the procedure

	Descending aorta			Aortic arch	RCCA	
	Displacement (mm)	Number twists	Time (s)	Torque-ccw (N.mm)	Speed (mm/s)	Displacement (mm)
Experienced	32.2	29	25	1.36	2.6	90.2
Inexperienced	20.3	70	37	0.48	5.5	82.5

Fig. 3 shows the result for force, torque and displacement of an experienced vs. inexperienced operator over the whole procedure. Distinct patterns can be detected, especially over the more complex parts of the anatomy (aortic arch and RCCA). Experienced operators rely on torque and small forces for maneuvering through these areas, and maximum torque is applied when transitioning from the arch to the artery. Overall forces applied by the experienced user are smaller and more uniform, and large forces are only used for specific controlled maneuvers such as advancing the catheter up the aorta or entering the aortic arch. Catheter displacements for experienced operators are also much smoother and contain less back and forth movement, while depicting difficulties for inexperienced operators when entering the arch as well as forcing the catheter from the arch into the narrow artery.

The similarity cost values obtained from dynamic time warping between the experienced and inexperienced operators for the three sections of the procedure
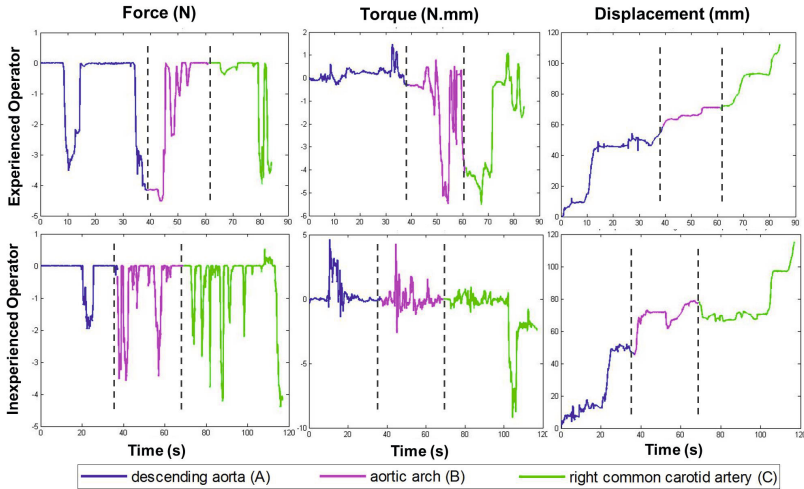
Fig. 3. Force, torque and displacement signals for experienced vs. inexperienced operator, with different colours showing the different sections of the procedure

show large differences in displacement in the first and last part of the procedure (1.55e3, 1.27e3, 1.92e3), highlighting the movement efficiency of experienced operators. Differences in torque are higher in the aortic arch and the carotid artery sections (441.18, 563.09, 894.29). There are also high differences between pull forces in the first and last part of the procedure (747.28, 351.32, 671.94), related to the back and forth movement of inexperienced operators. These findings directly map with the factor analysis results presented above.

The maximum force values (on average) in the axial direction over all procedures, for each of the three sections are 2.09 N, 3.03 N, and 2.93 N for experienced operators. The corresponding values for inexperienced operators are 2.86 N, 3.16 N, and 2.80 N. The maximum torque values for experienced operators over these sections are 2.84 N.mm, 5.26 N.mm, and 6.05 N.mm as compared to 1.87 N.mm, 2.00 N.mm and 6.71 N.mm for inexperienced users.

4   Discussion and Conclusion

A novel miniaturized proximal sensing platform is proposed to non-intrusively measure forces and torques applied during endovascular procedures and provide information on catheter dynamics and force and motion patterns used by operators over different levels of experience. A user study was performed to relate catheter tip motion to forces applied at the proximal end over different steps of a typical endovascular procedure. Different performance metrics related to catheter dynamics, proximal forces and torques, manipulation skills, and procedure time were compared for two groups of operators and patterns of correlation were extracted. It should be noted that the forces applied by the operator include forces to overcome friction from the introducer sheath as well as the vasculature.

The results highlight the importance of underlying factors and experience related skills that affect the efficiency, success and ergonomics of catheterization. Understanding these can also improve assessment and training of catheterization skills. The outcome of this research provides important insight into the perceptual cues used for optimized design of robotic catheter navigation systems while maintaining natural operator skills required for conventional catheter navigation.

References

1. Riga, C.V., Bicknell, C.D., Hamady, M.S., Cheshire, N.J.W.: Evaluation of robotic endovascular catheters for arch vessel cannulation. J. Vasc. Surg. 54(3), 799–809 (2011)
2. Saliba, W., Reddy, V.Y., Wazni, O., Cummings, J.E., Burkhardt, J.D., Haissaguerre, M., Kautzner, J., Peichl, P., Neuzil, P., Schibgilla, V., Noelker, G., Brachmann, J., Biase, L.D., Barrett, C., Jais, P., Natale, A.: Atrial fibrillation ablation using a robotic catheter remote control system: Initial human experience and long-term follow-up results. J. Am. Coll. Cardiol. 51(25), 2407–2411 (2008)
3. Riga, C., Bicknell, C., Cheshire, N., Hamady, M.: Initial clinical application of a robotically steerable catheter system in endovascular aneurysm repair. J. Endovasc. Ther. 16(2), 149–153 (2009)
4. Jayender, J., Patel, R.: Wave variables based bilateral teleoperation of an active catheter. In: 2nd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 27–32 (2008)
5. Jayender, J., Patel, R., Nikumb, S.: Robot-assisted catheter insertion using hybrid impedance control. In: IEEE International Conference on Robotics and Automation, pp. 607–612 (2006)
6. Marcelli, E., Cercenelli, L., Plicchi, G.: A novel telerobotic system to remotely navigate standard electrophysiology catheters. In: Computers in Cardiology, pp. 137–140 (2008)
7. Srimathveeravalli, G., Kesavadas, T., Li, X.: Design and fabrication of a robotic mechanism for remote steering and positioning of interventional devices. Int. J. Med. Robot. Comp. 6(2), 160–170 (2010)
8. Lin, P.H., Bush, R.L., Peden, E.K., Zhou, W., Guerrero, M., Henao, E.A., Kougias, P., Mohiuddin, I., Lumsden, A.B.: Carotid artery stenting with neuroprotection: assessing the learning curve and treatment outcome. Am. J. Surg. 190(6), 855–863 (2005)
9. Thakur, Y., Holdsworth, D.W., Drangova, M.: Characterization of catheter dynamics during percutaneous transluminal catheter procedures. IEEE Trans. Biomed. Eng. 56(8), 2140–2143 (2009)

Data-Driven Visual Tracking
in Retinal Microsurgery

Raphael Sznitman[1], Karim Ali[1], Rogério Richa[2], Russell H. Taylor[2],
Gregory D. Hager[2], and Pascal Fua[1]

[1] École Polytechnique Fédérale de Lausanne, Switzerland
[2] The Johns Hopkins University, Baltimore, USA
firstname.lastname@epfl.ch, {richa,rht,hager}@jhu.edu

Abstract. In the context of retinal microsurgery, visual tracking of in-
struments is a key component of robotics assistance. The difficulty of
the task and major reason why most existing strategies fail on *in-vivo*
image sequences lies in the fact that complex and severe changes in in-
strument appearance are challenging to model. This paper introduces a
novel approach, that is both data-driven and complementary to existing
tracking techniques. In particular, we show how to learn and integrate
an accurate detector with a simple gradient-based tracker within a ro-
bust pipeline which runs at framerate. In addition, we present a fully
annotated dataset of retinal instruments in *in-vivo* surgeries, which we
use to quantitatively validate our approach. We also demonstrate an
application of our method in a laparascopy image sequence.

1 Introduction

Retinal microsurgery (RM) is one of the few available treatments options for
many blinding eye conditions. During surgery, the operating surgeon uses a stereo
microscope to visualize the retina and manipulates a set of surgical instruments
(*i.e.* tipped forceps or picks) to perform the procedure, as depicted in Fig. 1.

Given its importance and the demanding nature of the surgery, a number
of new technologies have focused on improving aspects of RM. Some of these
technologies have included a steady-hand robot [1] or an instrument capable
of visualizing anatomical structures below the surface of the retina via optical
coherence tomography [2]. Yet, for these technologies to fully develop and ulti-
mately be incorporated into clinical environments, one missing component is the
ability to accurately and reliably estimate the location of an instrument when
in the camera field of view. With this in mind, this paper focuses on real-time
visual tracking of instruments in *in-vivo* RM monocular image sequences.

A major difficulty with this task is that instrument appearance is difficult to
model well over time. Most existing methods have relied instead on knowing the
instrument geometry beforehand to solve complex optimization problems [3,4], or
have constructed sophisticated and robust objective functions within more tra-
ditional gradient-based frameworks to deal with appearance change [5,6]. Typ-
ically, these methods have extremely simple appearance models that combine

geometry with colour or edge-based features, and ultimately work well only in limited conditions such as in eye phantoms. For example, using the method of [6], tracking is often lost after only 5 frames in the *in-vivo* sequence of Fig. 1. Note that a similar observation can also be made regarding tool tracking techniques for laparoscopic surgery [7].

In short visual tracking of instruments in *in-vivo* RM is characterized by complex appearance changes that existing approaches fail to handle. In contrast, this paper introduces an alternative approach, one that is data-driven and complementary to the aforementioned methods. In particular, we show how to integrate the framework of [8], which constructs accurate classifiers, for the task of instrument detection. Coupled with simple gradient-based tracking, our pipeline is extremely robust and runs at video framerate. In addition, we present a fully annotated dataset of retinal instruments in human *in-vivo* surgeries and quantitatively validate our pipeline on this dataset. Finally, we also demonstrate how our approach performs on a laparoscopy image sequence.

The remainder of this paper is organized as follows: We begin by describing our pipeline and its components in Sec. 2. In Sec. 3 we validate our method experimentally and conclude with final remarks in Sec. 4.

2   Method

To motivate our approach and pipeline, we begin with the following observations:

1. To work reliably, gradient-based trackers [6,9] need continuous template updating to maintain accurate position estimation when changes in the target appearance are severe.
2. Using reasonable amounts of training data (*e.g.* 500 positive examples), classifiers as in [8] provide excellent methods to detect the 2D location of a deformable target irrespective of its orientation.
3. Given that tracking is a sequential estimation problem, detection of targets can be restricted to promising locations provided by fast and moderately accurate methods.

Based on these observations, we propose a detection based scheme to track the 2D instrument tip position in *in-vivo* RM image sequences. Once initialized, our pipeline operates as follows: we first use a gradient based tracker to provide an approximate estimate of the target's new location. We then exhaustively evaluate a detector to predict the presence of an instrument in a reduced region of the image space, which is parametrized by tracker's estimate from the previous step. Finally, we use spatial and score weighting of the detector responses to provide accurate instrument position, and update the tracker template. This process is depicted in Fig. 1, and the following sections describe each component in detail. Note that, initialization of the instrument position, and reinitialization when the instrument is not found, is achieved by using the constructed detector and hence no user input is required in our pipeline.
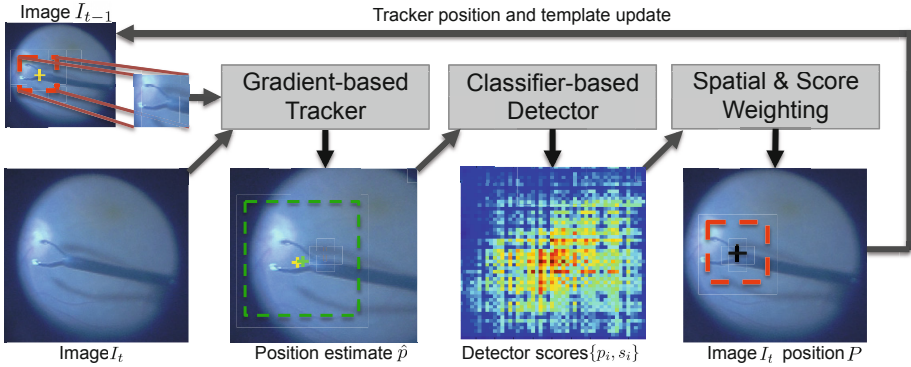
Fig. 1. Pipeline Diagram: First, an updated template and the previous frame instrument position (yellow cross) are used to initialize a gradient based tracker. The new position estimate, $\hat{p}$ (green cross) serves as the center of the region of interest (green box), that the detector evaluates at every location. At each location, positions and scores $\{p_i, s_i\}$ are computed and weighted to provide the final instrument position, P (black cross).

2.1   Tracking

In order to provide an approximate location for the instrument position, we first compute the displacement of a window centered at the previous tool tip location using a gradient-based tracking method. The method is based on Efficient Second-Order Minimization [9]. Assuming that no large illumination variations occur between sequential images, SSD was adopted as similarity measure. The reference template used in this step is updated at every new image using the tool tip position estimated from the previous image. In our experiments, we maintain a fixed template size: 50×50 pixels. This process results in a tool tip estimate, which we denote as $\hat{p}$. Fig. 1 shows an example of $\hat{p}$ (green cross) on a given frame. Note that alternative similarity measures could be substituted instead.

2.2   Detector

The strong appearance changes of tools during RM severely complicate the detection task. Standard learning based detection methods can only cope with deformations and rotations via a detailed labeling of training data and an exhaustive exploration of these parameters when evaluating the classifier. This latter point makes detection of targets particularly slow, which helps explain their lack of use so far. Recently however, a framework was presented in [8] which overcomes these difficulties. The authors design a set of so-called pose-estimator features which modulate feature extraction according to various image cues. The result is a deformable detector which can learn the deformations and rotations present in the training data. The method, based on AdaBoost, does not require an exploration of the pose parameters at test time and is thus well-suited for the task at hand.

We therefore use this framework along with the proposed set of deformable features, which compute sums of oriented edges in various image areas. To train this detector, positive and negative examples must be provided (*i.e.* instrument and non-instrument images, respectively). Here, we use square bounding boxes of the instruments indicating the location and spatial extent of the instrument for positive samples. Negative samples are randomly selected from the remainder of the images. One additional difficulty, not considered in [8], is how to efficiently train such a deformable detector of fixed size $r \times r$ from an image sequence exhibiting multi-scale data. To this end, we compute a Gaussian Pyramid for each image by successive smoothing and downsampling. For each positive example, its bounding box is replaced by an appropriately located box of size $r \times r$ at the Gaussian Pyramid level l which results in the best $r \times r$ approximation of the original sample. Detection proceeds in a similar fashion with each image decomposed into a Gaussian Pyramid and our $r \times r$ detector exhaustively visiting every location in the Pyramid.

Given the approximate instrument position provided by the tracker, we only evaluate the classifier at each location in a 50×50 region of the image, centered on the position estimate provided by the tracker. This results in a set of pixel positions and associated unsigned classifier score, $\{p_i, s_i\}$.

2.3 Estimating Instrument Position

Given the position estimate of the tracker, $\hat{p}$ and the set of detection scores $\{p_i, s_i\}$, we now describe how to combine these estimates to provide the final instrument location.

We first perform a weighting of the classifier scores with regard to their spatial placement. In particular, we favour locations that are near the position estimate provided by the tracker. That is, we first compute spatially adjusted scores, $\tilde{s}$, $\tilde{s}_i = s_i e^{-\frac{1}{2\sigma^2}(p_i - \hat{p})^2}$, where σ is half the radius of the search window ($\sigma = \frac{50}{2} = 25$ in our experiments). Then, instead of doing non-maximum suppression as in [8], we estimate the final position of the instrument, P, by averaging the weighted scores, $\tilde{s}_i$, $P = \frac{\sum_i^N \tilde{s}_i p_i}{\sum_i^N \tilde{s}_i}$. This effectively reduces the effect of extreme scores and outlier influence by weighted voting. We consider a detection valid if the score associated with the location P is above a threshold (in practice it is set to provide a 80% true positive rate). When no instrument is found in a frame, then detector is then evaluated at all locations of subsequent images, until a new instrument location is found.

3 Experiments and Results

The presented pipeline is implemented in C++, and all experiments were performed on a MacBook Pro, 2.5 GHz Quad core computer with 4GB RAM. Our pipeline runs at 15fps and should run even faster implemented on a GPU.
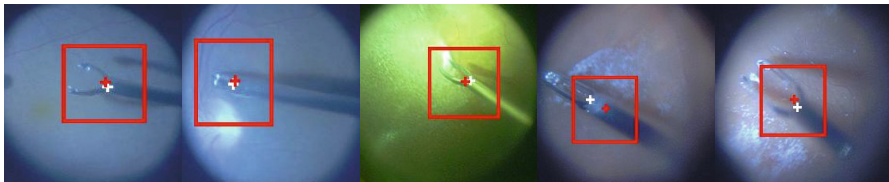
Fig. 2. Example of instrument detections (red) in our *in-vivo* dataset with annotated ground truth (white cross). The pixel distance error for each example are: 8, 7, 11, 25 and 18. See Video 1 for full detection and tracking sequence.

3.1   Retina Microsurgery Dataset

We begin by introducing a fully annoted dataset for RM instrument detection and tracking. This image set consists of 4 sequences of *in-vivo* vitreoretinal surgery, containing a total of 1500 images (640 × 480 pixels). Fig. 2 shows representative images from the dataset, illustrating variations in illumination type and quantity, light source position and the presence of blur and shadows. Different types of cameras were used to acquire the images but in each case the video was collected directly from the surgical microscope. Calibration data was not available since the surgeon frequently varies the focal length during the procedure. Each image contains at most one instance of a tool, with some images being tool free. The tool tip of each instrument has been annotated by hand. This dataset is publicly available online via the corresponding authors website, at https://sites.google.com/site/sznitr/.

Full Dataset Evaluation. In our first experiment, we trained our classifier by using the first half of each sequence in the above dataset and evaluated our method on the remaining sequence halves. The result of our pipeline can be seen in Video 1, with some snapshots shown in Fig. 2 (see above website for associated videos). In general, consistent tracking is achieved even in cases of strong appearance changes.

To provide some quantitative validation of our method we plotted the proportion of frames where the instrument tip was determined correctly, as function of sensitivity of the detection criteria. More specifically, we defined a correct detection to be any pixel estimation that is within δ pixels of the groundtruth annotation. Fig. 3(left) show this plot when varying δ between 15 and 40. 15 pixel may appear as a large starting threshold, but consider that the average tool shaft diameter in the dataset is of 20 pixels, and due to blurring and illumination changes throughout the sequences, the annotations themselves are noisy (see Fig. 2). Hence, smaller threshold results are not particularly meaningful here.

We compare our approach to three existing gradient based trackers on the same set of images: the Mutual Information of [6], the SCV of [10] and the SSD tracker used in this pipeline. To allow a fair comparison, when any of these trackers provided false detections they were re-initialized with the ground truth,
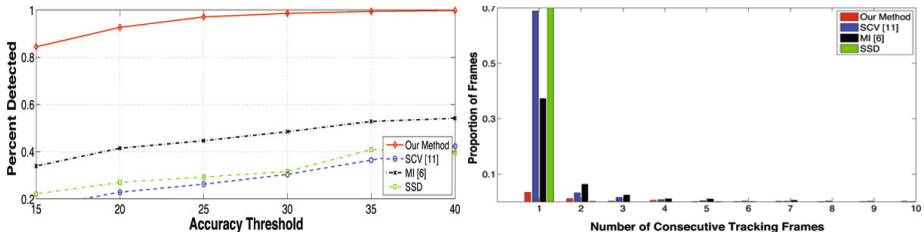
Fig. 3. Tracking Accuracy. *(left)* we show the percent of correctly detected instruments as function of the accuracy threshold. In red, our approach clearly outperforms state-of-the-art gradient-based trackers. *(right)* Proportion of frames for each number of consecutive correct detections.

and we report the proportion of frames where re-initialization is not required. From the figures, we clearly see that our approach outperforms that all three trackers. For example, for $\delta = 20$ our approach detects over 70% more than [6] and over 40% more than [10]. This corresponds to 449 and 309 more correct instrument detections, respectively. We also show in Fig. 3(right) the proportion of time where a certain number of consecutive correct detections ($\delta = 20$) took place. In particular, we see that the SCV and MI can only track for 1 frame over 35% and 65% of the time, while this only occurs 11% of the time for our approach. On average our method tracks for 25 consecutive frames while the SCV and MI achieve 2 and 5 frames, respectively. Also, in the cases where our method did lose tracking, correct reinitialization occurred on average after 1.5 frames.

Generalization: Detection-based methods as this one are often criticized for needing large amounts of training data and only working well on images similar to those found in the training set. To demonstrate, that this can be avoided, we show that even when training our classifier on three sequences, and testing on an unseen fourth, reliable tracking is achieved. As in typical cross-validation protocols, we trained our classifier on 3 sequences, and tested on the remaining set. We did this for 3 different sets (the 4th set was not usable in this case, since it contains no instruments in it). Fig. 4 shows training and testing image examples for each experiment (*i.e.* Exp.2a through c). Videos 2a though 2c show the tested sequences for these experiments. As in the previous case, we plotted detection accuracy against the detection criteria, showing that our method significantly outperforms [6] and [10] on all three sequences.

3.2 Laparoscopy Sequence

Finally, we briefly show how our approach can work for laparoscopic instrument tracking. Here, we downloaded a video sequence from Youtube[1], extracted images and hand labeled the locations of instruments in 1000 images. This provided roughly 2000 instrument locations (two instruments per image). From this, we

---

[1] http://www.youtube.com/watch?v=IVp1sgjQ5To
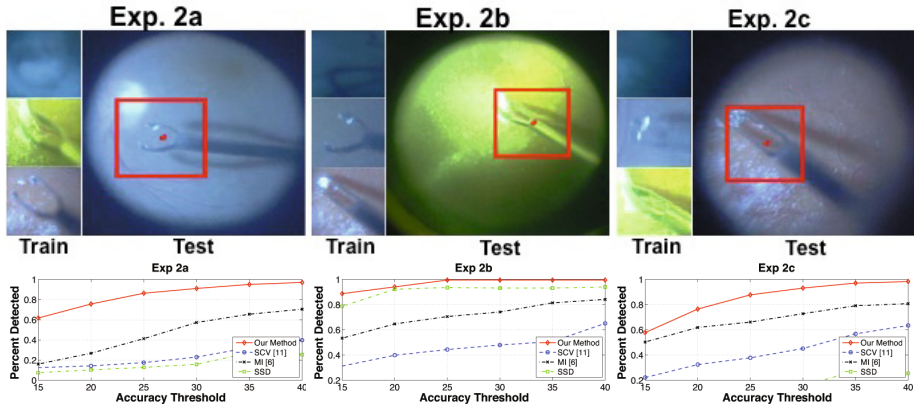
Fig. 4. Generalization experiments by training on 3 image sequences and tested on an unseen fourth sequence. Accuracy plots are also shown for each experiment.

trained our classifier on the first 500 images and evaluated our pipeline on the remaining images. Given, that two instruments are present in frames, we proceeded as in the RM case, found an instrument, suppressed it from the image, and repeated this process for the second tool. Otherwise, the pipeline is identical to that of the previous experiments.

Video 5 shows the result of our pipeline, of which a few frames are shown in Fig. 5. In summary, tracking is maintained for a substantial number of frames. However, two main failing points can be seen: 1) Extreme changes in instrument structure, that were not observed in the training sequences, are poorly handled by our system (as shown in Fig. 5(right)), 2) occluded instruments are not found given that there is no geometrical model to help with such situations. A possible alternative to overcome this may be to integrate our approach with more elaborate prior instrument knowledge (as in [3,4,7]).
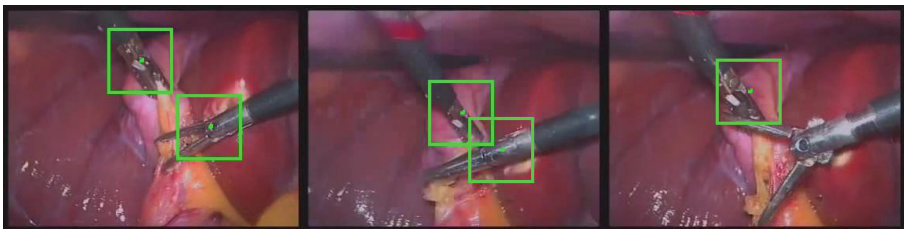
Fig. 5. Example of our approach tracking two instruments during Laparoscopic surgery

4   Conclusion

We presented an alternative approach for visual detecting and tracking retinal instruments during *in-vivo* retinal microsurgery. Our technique involves training

a highly accurate instrument detector, coupled with a simple gradient based tracker to produce reliable tracking. Soft weighting of both classifier scores and locations are fused to produce accurate position estimates even in challenging cases. We extensively validated our method on a fully annotated *in-vivo* dataset, where we showed consistent tracking. We also demonstrated the applicability of our approach on a laparoscopy image sequence.

References

1. Uneri, A., Balicki, M., Handa, J., Gehlbach, P., Taylor, R., Iordachita, I.: New steady-hand eye robot with micro-force sensing for vitreoretinal surgery. In: 3rd IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 814–819 (September 2010)
2. Balicki, M., Han, J.H., Iordachita, I., Gehlbach, P., Handa, J., Taylor, R., Kang, J.: Single Fiber Optical Coherence Tomography Microsurgical Instruments for Computer and Robot-Assisted Retinal Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 108–115. Springer, Heidelberg (2009)
3. Pezzementi, Z., Voros, S., Hager, G.D.: Articulated object tracking by rendering consistent appearance parts. In: IEEE International Conference on Robotics and Automation, pp. 3940–3947 (May 2009)
4. Sznitman, R., Basu, A., Richa, R., Handa, J., Gehlbach, P., Taylor, R., Jedynak, B., Hager, G.: Unified Detection and Tracking in Retinal Microsurgery. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 1–8. Springer, Heidelberg (2011)
5. Burschka, D., Corso, J.J., Dewan, M., Lau, W., Li, M., Lin, H., Marayong, P., Ramey, N., Hager, G.D., Hoffman, B., Larkin, D., Hasser, C.: Navigating inner space: 3-d assistance for minimally invasive surgery. Robotics and Autonomous Systems 52, 5–26 (2005)
6. Richa, R., Balicki, M., Meisner, E., Sznitman, R., Taylor, R., Hager, G.: Visual Tracking of Surgical Tools for Proximity Detection in Retinal Surgery. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 55–66. Springer, Heidelberg (2011)
7. Voros, S., Long, J.A., Cinquin, P.: Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders. International Journal of Robotic Research 26(11-12), 1173–1190 (2007)
8. Ali, K., Fleuret, F., Hasler, D., Fua, P.: A real-time deformable detector. Transactions on Pattern Analysis and Machine Intelligence 34(2), 225–239 (2011)
9. Benhimane, S., Malis, E.: Homography-based 2D Visual Tracking and Servoing. International Journal of Robotics Research 26(7), 661–676 (2007)
10. Pickering, M., Muhit, A., Scarvell, J., Smith, P.: A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In: Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, pp. 5821–5824 (September 2009)

Real-Time Motion Compensated Patient Positioning and Non-rigid Deformation Estimation Using 4-D Shape Priors

Jakob Wasza[1], Sebastian Bauer[1], and Joachim Hornegger[1,2]

[1] Pattern Recognition Lab, Department of Computer Science
[2] Erlangen Graduate School in Advanced Optical Technologies (SAOT)
Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
{jakob.wasza,sebastian.bauer,joachim.hornegger}@cs.fau.de

Abstract. Over the last years, range imaging (RI) techniques have been proposed for patient positioning and respiration analysis in motion compensation. Yet, current RI based approaches for patient positioning employ rigid-body transformations, thus neglecting free-form deformations induced by respiratory motion. Furthermore, RI based respiration analysis relies on non-rigid registration techniques with run-times of several seconds. In this paper we propose a real-time framework based on RI to perform respiratory motion compensated positioning and non-rigid surface deformation estimation in a joint manner. The core of our method are pre-procedurally obtained 4-D shape priors that drive the intra-procedural alignment of the patient to the reference state, simultaneously yielding a rigid-body table transformation and a free-form deformation accounting for respiratory motion. We show that our method outperforms conventional alignment strategies by a factor of 3.0 and 2.3 in the rotation and translation accuracy, respectively. Using a GPU based implementation, we achieve run-times of 40 ms.

1 Introduction

Accurate patient positioning and respiratory motion analysis are key issues for the success of medical procedures. For example, in fractionated radiotherapy, the patient must be aligned with respect to planning data and continuously monitored to account for respiratory motion and spontaneous movements [1]. Recently, techniques based on real-time range imaging (RI) have been proposed for positioning and respiration analysis [2–6]. Compared to commonly employed imaging techniques, RI sensors are marker-less, non-intrusive and do not imply radiation exposure. However, current RI based positioning systems employ rigid registration techniques neglecting free-form deformations induced by respiratory motion. In this context, recent work reported an error scale of non-rigid motion up to 25 mm [5],[6]. Another motivation for non-rigid registration of respiration-induced surface deformations arises from observations that an analysis of multiple surface regions allows for an improved prediction of internal organ movement compared to a single external surrogate [7]. Apart from applications

in radiotherapy, the prediction of internal movement from external surrogates also holds potential for statistical lung motion modelling [8] or 4-D CT reconstruction and sorting [9]. However, current surrogates based on RI techniques rely on heuristic surface partitioning strategies based on points and patches [7] or manually selected bounding boxes of anatomical regions [4]. Though dense non-rigid surface registration techniques to generate respiration surrogates were recently proposed [6],[10], run-times of several seconds are not acceptable in clinical practice.

In this paper we propose a framework that by design allows to perform respiratory motion compensated patient positioning and non-rigid surface deformation estimation in a joint manner. Our method is based on pre-procedurally obtained patient-specific 4-D shape priors derived from statistical analysis of nonrigidly registered 3-D RI patient surface data from different respiration states. The intra-procedural alignment of the patient to the reference state is then driven by the shape priors, simultaneously yielding a rigid-body table transform and a free-form deformation accounting for respiratory motion. With real-time constraints in mind, our framework outsources the computationally expensive task of non-rigid surface registration to a pre-procedural training phase. Intraprocedurally, this allows to employ real-time algorithms taking advantage of the pre-procedurally obtained priors. In particular, we designed our method to support parallel computation on many-core architectures such as GPUs.

2 Method

Our method relies on RI devices that deliver dense and metric surface information of the captured body in real-time. Clinically available systems are for example AlignRT (Vision RT, London, UK) or Catalyst (C-RAD AB, Uppsala, Sweden). We denote $\boldsymbol{x}\left(\boldsymbol{\xi}_i\right) \in \mathbb{R}^3$ the world coordinate associated with $\boldsymbol{\xi}_i$ in a 2-D sampling domain $\Omega \subset \mathbb{R}^2$ discretized with $N \times M$ pixels. We denote a point cloud or surface as $\mathcal{S} = \{\boldsymbol{x}\left(\boldsymbol{\xi}_1\right), \ldots, \boldsymbol{x}\left(\boldsymbol{\xi}_{N \cdot M}\right)\}$ which can be linearized as:

$$\mathcal{S} \equiv \boldsymbol{s} = \left[\boldsymbol{x}\left(\boldsymbol{\xi}_1\right)^\top, \ldots, \boldsymbol{x}\left(\boldsymbol{\xi}_{N \cdot M}\right)^\top\right]^\top \in \mathbb{R}^{3NM}. \tag{1}$$

We now briefly explain the general steps of our method. Pre-procedurally, we non-rigidly register RI surfaces $\{\mathcal{S}_1, \ldots, \mathcal{S}_T\}$ acquired at different respiration states $t = 1 \ldots T$ to a reference surface $\mathcal{S}_{\mathrm{Ref}}$. The reference can be acquired using RI sensors or imported from volumetric planning data by (i) segmenting the body from the background, (ii) isosurface extraction, and (iii) rendering the surface mesh to a z-buffer representation and sample it from Ω. We obtain T displacement fields $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T\}$, $\boldsymbol{u}_t \in \mathbb{R}^{3NM}$ that are used to build a deformable model $\mathcal{M}\left(\boldsymbol{b}\right)$ representing the priors of our method. Here, $\boldsymbol{b}$ denotes a parameter vector to control the model. The intra-procedural respiratory motion compensated alignment of the patient to the reference state is then computed by finding a rigid-body transformation $(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{t}})$ with rotation $\widehat{\boldsymbol{R}}$ and translation $\widehat{\boldsymbol{t}}$ and the parameters $\widehat{\boldsymbol{b}}$ such that the corresponding model instance $\mathcal{M}(\widehat{\boldsymbol{b}})$ fits the patients instantaneous state. An illustration is given in Fig. 1.
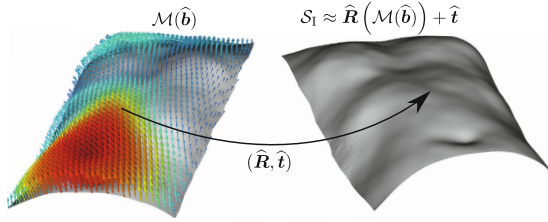
Fig. 1. Respiratory motion compensated patient alignment. The reference state (left gray shape) is deformed according to $\mathcal{M}(\widehat{\boldsymbol{b}})$ and transformed by $(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{t}})$ to fit the instantaneous state $\mathcal{S}_\mathrm{I}$ (right shape). The displacement field corresponding to $\mathcal{M}(\widehat{\boldsymbol{b}})$ is color coded with red tones denoting large and blue tones denoting small displacements.

2.1   Non-rigid Surface Registration

Training a deformable surface model requires a set of displacement fields $\mathcal{U}$ that describe the elastic deformation ϕ_t matching a patient reference surface $\mathcal{S}_\mathrm{Ref}$ and patient surface data $\mathcal{S}_t$ captured at different respiration states t. We represent the deformation ϕ_t by a displacement field $\boldsymbol{u}_t$ defined on Ω with

$$\phi_t\left(\boldsymbol{x}\left(\boldsymbol{\xi}\right)\right) = \boldsymbol{x}(\boldsymbol{\xi}) + \boldsymbol{u}_t(\boldsymbol{\xi}), \ \boldsymbol{x}\left(\boldsymbol{\xi}\right) \in \mathcal{S}_\mathrm{Ref}. \tag{2}$$

Now, the goal is to estimate ϕ_t in a sense that $\phi_t(\mathcal{S}_\mathrm{Ref}) \approx \mathcal{S}_t$. For this purpose, we represent $\mathcal{S}_t$ at time t by its corresponding signed distance function $d_t(\boldsymbol{x}) := \pm dist(\boldsymbol{x}, \mathcal{S}_t)$, where the sign is positive outside the body and negative inside. Further, $\nabla d_t(\boldsymbol{x})$ is the outward pointing normal on $\mathcal{S}_t$ and $|\nabla d_t(\boldsymbol{x})| = 1$. Based on $d_t(\boldsymbol{x})$, we can define the projection $P(\boldsymbol{x}) := \boldsymbol{x} - d_t(\boldsymbol{x})\nabla d_t(\boldsymbol{x})$ of a point $\boldsymbol{x} \in \mathbb{R}^3$ in a neighborhood of $\mathcal{S}_t$ onto the closest point on $\mathcal{S}_t$. Thus, we quantify the closeness of a displaced reference surface point $\phi_t(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_\mathrm{Ref}$ to $\mathcal{S}_t$ using $|P(\phi_t(\boldsymbol{x})) - \phi_t(\boldsymbol{x})| = |d_t(\phi_t(\boldsymbol{x}))|$ as a point-wise measure. Based on this closeness measure, we use a variational formulation for estimating $\boldsymbol{u}_t$ as a minimizer of the functional

$$\mathcal{E}[\boldsymbol{u}_t] = \mathcal{E}_\mathrm{match}[\boldsymbol{u}_t] + \alpha\mathcal{E}_\mathrm{reg}[\boldsymbol{u}_t] = \int_\Omega \left(d_t\big(\boldsymbol{x}(\boldsymbol{\xi}) + \boldsymbol{u}_t(\boldsymbol{\xi})\big)^2 + \alpha\|D\boldsymbol{u}_t(\boldsymbol{\xi})\|_F^2\right) d\boldsymbol{\xi} \tag{3}$$

where $D\boldsymbol{u}(\boldsymbol{\xi})$ denotes the Jacobian of $\boldsymbol{u}(\boldsymbol{\xi})$ and α the regularization weight. The matching term $\mathcal{E}_\mathrm{match}$ ensures that $\phi_t(\mathcal{S}_\mathrm{Ref}) \approx \mathcal{S}_t$. As a smoothness prior, we took a quadratic regularization term on the Jacobian of the displacement into account. For numerical minimization of Eq. 3, we considered a Finite Element approximation on a uniform rectangular $N \times M$ grid covering Ω for spatial discretization and applied a multi-scale gradient descent scheme, c.f. [10].

2.2   Deformable Model Generation

Based on the patient-specific set of non-rigidly registered surfaces $\mathcal{S}_t$ we apply a principal component analysis (PCA) to the training set $\mathcal{V} = \{\boldsymbol{s}_\mathrm{Ref}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_T\}$

with $\boldsymbol{v}_t = \boldsymbol{s}_{\mathrm{Ref}} + \boldsymbol{u}_t$. The $P \leq T$ eigenvectors $\boldsymbol{e}_i$ with the largest non-zero eigenvalues of the covariance matrix of the centered version of $\mathcal{V}$ define the model's modes of variation as

$$\boldsymbol{\Phi} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_P] \in \mathbb{R}^{3NM \times P}. \tag{4}$$

A linear combination of the P principal modes of variation spans a subset of RI surfaces $\mathcal{S}$ composed of the given modes of variation as

$$\mathcal{S} \equiv \mathcal{M}(\boldsymbol{b}) = \overline{\boldsymbol{v}} + \boldsymbol{\Phi}\boldsymbol{b}, \quad \overline{\boldsymbol{v}} = \frac{1}{T+1}\left(\boldsymbol{s}_{\mathrm{Ref}} + \sum_{t=1}^{T} \boldsymbol{v}_t\right). \tag{5}$$

Here, $\boldsymbol{b} \in \mathbb{R}^P$ is a parameter vector holding the coefficients for the modes of variation, thus accounting for the model's inherent free-form deformations.

2.3 Respiratory Motion Compensated Patient Alignment

The intra-procedural alignment of the patient is performed by fitting the deformable model $\mathcal{M}$ to the patient's instantaneous respiration state $\mathcal{S}_\mathrm{I}$:

$$\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{t}}, \widehat{\boldsymbol{b}} = \underset{\boldsymbol{R}, \boldsymbol{t}, \boldsymbol{b}}{\operatorname{argmin}} \; dist\left(\mathcal{S}_\mathrm{I}, \boldsymbol{R}\left(\overline{\boldsymbol{v}} + \boldsymbol{\Phi}\boldsymbol{b}\right) + \boldsymbol{t}\right). \tag{6}$$

The rotation matrix $\widehat{\boldsymbol{R}}$ and translation vector $\widehat{\boldsymbol{t}}$ define the rigid-body table transformation whereas the model parameter vector $\widehat{\boldsymbol{b}}$ accounts for non-rigid deformations induced by respiratory motion. Furthermore, $dist\left(\mathcal{S}_i, \mathcal{S}_j\right)$ quantifies the distance between two surfaces and may for example denote a point-to-point or point-to-plane measure. In this work we employ a model fitting strategy derived from the *iterative closest point* (ICP) algorithm [11]. In each iteration k, the current model instance $\mathcal{S}_k$ is aligned to the instantaneous surface $\mathcal{S}_\mathrm{I}$ via a rigid-body transformation $\left(\boldsymbol{R}_k, \boldsymbol{t}_k\right)$ estimated by a closest-point relationship. Based on the estimated transformation, the model's closest points $\boldsymbol{Y}_k$ on the instantaneous surface $\mathcal{S}_\mathrm{I}$ are then projected onto the model basis $\boldsymbol{\Phi}$ to update the parameters $\boldsymbol{b}_k$. See Algorithm 1 for details.

A benefit of this model fitting scheme is the inherent high degree of parallelism in each iteration of Algorithm 1, thus allowing for a real-time capable implementation on many-core systems such as GPUs. In particular, we employ the recently proposed *random ball cover* for efficient closest point search [12].

2.4 Non-rigid Deformation Estimation Using 4-D Shape Priors

By design, our method allows for efficient estimation of respiration induced deformations. For two surfaces $\left(\mathcal{S}_i, \mathcal{S}_j\right)$ and its associated estimated model coefficients $\left(\boldsymbol{b}_i, \boldsymbol{b}_j\right)$, the estimation of a dense displacement field $\boldsymbol{u}_{i,j}$ between these surfaces breaks down to a linear mapping as $\boldsymbol{u}_{i,j} = \boldsymbol{\Phi}\left(\boldsymbol{b}_i - \boldsymbol{b}_j\right)$. This follows directly from Eq. (5). We further note that the computational complexity of non-rigid deformation estimation is therefore decoupled from the actual non-rigid registration technique creating the priors contained in $\boldsymbol{\Phi}$. Inherently, this allows to employ computational expensive non-rigid registration techniques as proposed by Schaerer et al. [6] or Bauer et al. [10] in the model generation stage.

Algorithm 1. Algorithm for Model Fitting

  Input: Instantaneous surface $\mathcal{S}_{\mathrm{I}}$ and deformable model $\overline{\boldsymbol{v}}, \boldsymbol{\Phi}$
  Input: Initial guess for table transformation $\boldsymbol{R}_0, \boldsymbol{t}_0$ and model parameters $\boldsymbol{b}_0$
  for $k = 1, \ldots, N_{\mathrm{iter}}$ **do**
    $\mathcal{S}_k = \boldsymbol{R}_{k-1}\left(\overline{\boldsymbol{v}} + \boldsymbol{\Phi}\boldsymbol{b}_{k-1}\right) + \boldsymbol{t}_{k-1}$
    $\boldsymbol{Y}_k \Leftarrow \texttt{closestPoints}\left(\mathcal{S}_k, \mathcal{S}_{\mathrm{I}}\right)$
    $(\boldsymbol{R}_k, \boldsymbol{t}_k) \Leftarrow \texttt{estimateTransformation}\left(\mathcal{S}_k, \boldsymbol{Y}_k, \boldsymbol{R}_{k-1}, \boldsymbol{t}_{k-1}\right)$
    $\boldsymbol{b}_k = \boldsymbol{\Phi}^{\top}\left(\boldsymbol{R}_k^{-1}\left(\boldsymbol{Y}_k - \boldsymbol{t}_k\right) - \overline{\boldsymbol{v}}\right)$
  end for
  Output: Estimated table transformation $\left(\widehat{\boldsymbol{R}}, \widehat{\boldsymbol{t}}\right)$ and model coefficients $\widehat{\boldsymbol{b}}$

---

3   Evaluation and Results

For our experiments we captured RI data from six subjects using a structured light sensor (30 Hz, 640×480 px) in a clinical environment. RI data were enhanced using edge-preserving filtering, cropped to a region of interest covering the subject and re-sampled in the sampling domain Ω to a resolution of 128×128 pixels. For model generation, the subjects were instructed to perform thoracic and abdominal breathing subsequently. For each breathing mode we then extracted surface data $\mathcal{S}_t$ from $T = 8$ phases covering one respiration cycle. We found that more cycles did not improve accuracy. For all subjects, the body surface at full abdominal expiration was chosen as the reference $\mathcal{S}_{\mathrm{Ref}}$ for non-rigid registration of the remaining 15 shapes to form the training set $\mathcal{V}$. PCA was then performed on the training set $\mathcal{V}$ with the number of modes P chosen such that 99% of the input variance is explained. This resulted in four variation modes for all subjects. To validate the extrapolation ability of our model to unseen data, the subjects were asked to perform regular breathing over several respiration cycles. Surfaces for validation were generated from 32 RI frames sampled every 10th frame, starting at an arbitrary phase.

Patient Positioning. The accuracy of our method for respiratory motion compensated patient positioning is assessed by evaluating the difference between the estimated table transforms $(\boldsymbol{R}_i, \boldsymbol{t}_i)$ for validation frame i and the corresponding ground truth (GT) transform $\boldsymbol{R}_{\mathrm{GT}} \equiv \left\{r_{\mathrm{GT}}^1, r_{\mathrm{GT}}^2, r_{\mathrm{GT}}^3\right\}$ and $\boldsymbol{t}_{\mathrm{GT}} \equiv \left\{t_{\mathrm{GT}}^1, t_{\mathrm{GT}}^2, t_{\mathrm{GT}}^3\right\}$. The superscripts $1, 2, 3$ denote rotation around or translation along the x, y and z axis, respectively. The GT transform was derived from the identity transform as the subjects did not change position between the training and testing phase. For each subject, the average rotation error Δr_i and translation error Δt_i for validation frame i is computed according to:

$$\Delta r_i = \frac{1}{3}\sum_{j=1}^{3}\left|r_{\mathrm{GT}}^j - r_i^j\right|, \quad \Delta t_i = \frac{1}{3}\sum_{j=1}^{3}\left|t_{\mathrm{GT}}^j - t_i^j\right|. \tag{7}$$

For comparison, we oppose our method to a conventional ICP-based rigid alignment strategy as used in [5]. For both methods, the initial estimate for the table
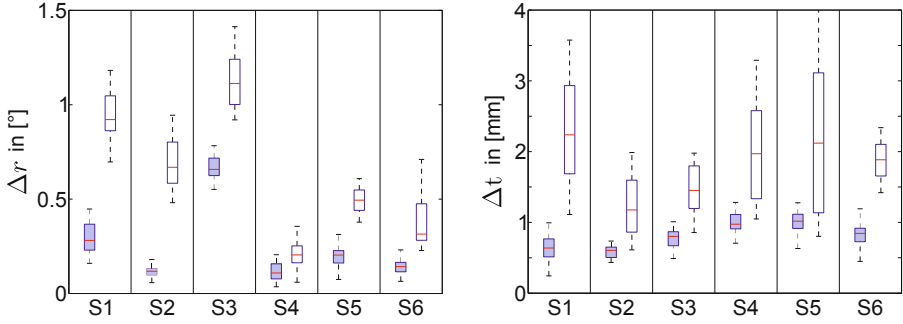
Fig. 2. Table transform error of our method (shaded) compared to conventional ICP-based alignment (not shaded) for subjects S1-S6. The left figure depicts the mean rotation error Δr in [°], the right figure the mean translation error Δt in [mm] compared to the ground truth transform.

transform was set to 10 mm and 5° off from the GT position. Fig. 2 shows quantitative results for the average rotation and translation error over all 32 validation frames. Note that our motion compensated positioning method significantly reduces that table transform error. Over all subjects, the rotation and translation error decreases by a factor of 3.0 and 2.3, respectively.

Deformation Estimation. The capability of our method for non-rigid deformation estimation is assessed by computing the absolute error in terms of surface mismatch between the transformed estimated model instance and the instantaneous surface as $|dist\left(\boldsymbol{R}(\mathcal{M}\left(\boldsymbol{b}\right)) + \boldsymbol{t}, \mathcal{S}_{\mathrm{I}}\right)|$ using a point-to-plane measure. In order to account for boundary effects such as edges, we restricted the evaluation to a center region on the subjects surface. For comparison, we again oppose our method to a conventional ICP-based rigid alignment strategy. Fig. 3 exemplarily illustrates the surface mismatch for subject S1. In Fig. 4 left, the surface mismatch is shown for subject S4 and a subset of validation frames. The right plot in Fig. 4 depicts the surface mismatch for the individual subjects over all validation frames. We note the large residual displacements with ICP-based alignment that can be reduced significantly with our model-based deformation estimation. Over all subjects, our method reduces the surface mismatch by a factor of 1.9.

Performance Evaluation. We implemented the model fitting routine as described in Algorithm 1 on an NVIDIA GTX570 GPU using the CUDA architecture. For performance reasons, we employed a sparse fitting scheme, e.g. a subset of uniformly sampled points in the model $\mathcal{M}$ as well as the instantaneous surface $\mathcal{S}_{\mathrm{I}}$ where used. For $4 \cdot 10^3$ model points, $1.6 \cdot 10^4$ surface points and $N = 100$ iterations that proved to be sufficient to achieve the accuracy reported above, our method runs at 40 ms.
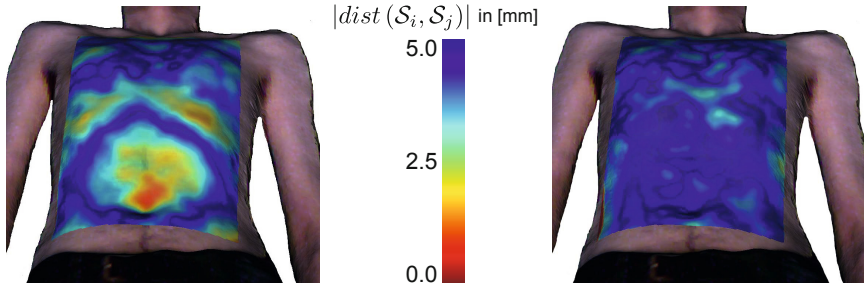
Fig. 3. Color-coded surface mismatch for ICP-based alignment (left) and our method (right). Average misalignment error is 1.3 mm with ICP and 0.5 mm with our method.
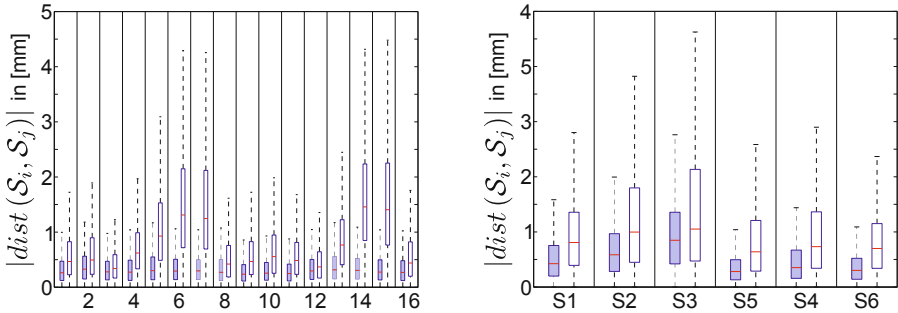
Fig. 4. Surface mismatch for our method (shaded) and ICP-based alignment (not shaded). The left figure shows the surface mismatch for an individual subject over several frames. The right figure depicts the surface mismatch for the individual subjects over all frames.

4   Conclusion

In this paper we have proposed a real-time capable range imaging based framework for joint respiratory motion compensated patient positioning and non-rigid surface deformation estimation. Both tasks are achieved by the main contribution of this paper, namely to employ 4-D patient-specific shape priors obtained from statistical analysis of non-rigidly registered surfaces. In experiments we showed that, in comparison to conventional positioning strategies, our method reduces the average alignment error by a factor of 3.0 and 2.3 for the rotational and translational components, respectively. We further demonstrated that a GPU-based implementation of our method allows to estimate dense surface deformations at a run-time of 40 ms.

References

1. Bert, C., Durante, M.: Motion in radiotherapy: particle therapy. Phys. Med. Biol. 56(16), R113–R144 (2011)
2. Schaller, C., Rohkohl, C., Penne, J., Stürmer, M., Hornegger, J.: Inverse C-arm Positioning for Interventional Procedures Using Real-Time Body Part Detection. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 549–556. Springer, Heidelberg (2009)
3. Fayad, H., Pan, T., Roux, C., Le Rest, C., Pradier, O., Clement, J., Visvikis, D.: A patient specific respiratory model based on 4D CT data and a time of flight camera (TOF). In: IEEE Nuclear Science Symposium Conference Record (NSS/MIC), pp. 2594–2598 (2009)
4. Hughes, S., McClelland, J., Tarte, S., Lawrence, D., Ahmad, S., Hawkes, D., Landau, D.: Assessment of two novel ventilatory surrogates for use in the delivery of gated/tracked radiotherapy for non-small cell lung cancer. Radiother. Oncol. 91(3), 336–341 (2009)
5. Placht, S., Stancanello, J., Schaller, C., Balda, M., Angelopoulou, E.: Fast time-of-flight camera based surface registration for radiotherapy patient positioning. Med. Phys. 39(1), 4–17 (2012)
6. Schaerer, J., Fassi, A., Riboldi, M., Cerveri, P., Baroni, G., Sarrut, D.: Multi-dimensional respiratory motion tracking from markerless optical surface imaging based on deformable mesh registration. Phys. Med. Biol. 57(2), 357–373 (2012)
7. Fayad, H., Pan, T., Clement, J.F., Visvikis, D.: Technical note: Correlation of respiratory motion between external patient surface and internal anatomical landmarks. Med. Phys. 38(6), 3157–3164 (2011)
8. Klinder, T., Lorenz, C., Ostermann, J.: Prediction Framework for Statistical Respiratory Motion Modeling. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 327–334. Springer, Heidelberg (2010)
9. Gianoli, C., Riboldi, M., Spadea, M.F., Travaini, L.L., Ferrari, M., Mei, R., Orecchia, R., Baroni, G.: A multiple points method for 4D CT image sorting. Med. Phys. 38(2), 656–667 (2011)
10. Bauer, S., Berkels, B., Hornegger, J., Rumpf, M.: Joint ToF Image Denoising and Registration with a CT Surface in Radiation Therapy. In: Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M. (eds.) SSVM 2011. LNCS, vol. 6667, pp. 98–109. Springer, Heidelberg (2012)
11. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2), 239–256 (1992)
12. Cayton, L.: Accelerating nearest neighbor search on manycore systems. Computing Research Repository abs/1103.2635 (2011)

Semi-automatic Catheter Reconstruction
from Two Views

Matthias Hoffmann[1,*], Alexander Brost[1], Carolin Jakob[1], Felix Bourier[2],
Martin Koch[1], Klaus Kurzidim[2], Joachim Hornegger[1], and Norbert Strobel[3]

[1] Pattern Recognition Lab.,
Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
`Matthias.Hoffmann@cs.fau.de`
[2] Krankenhaus Barmherzige Brüder, Regensburg, Germany
[3] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. We propose novel methods for (a) detection of a catheter
in fluoroscopic images and (b) reconstruction of this catheter from two
views. The novelty of (a) is a reduced user interaction and a higher
accuracy. It requires only a single seed point on the catheter in the fluo-
roscopic image. Using this starting point, possible parts of the catheter
are detected using a graph search. An evaluation of the detection using
66 clinical fluoroscopic images yielded an average error of 0.7 mm ± 2.0
mm. The novelty of (b) is a better ability to deal with highly curved
objects as it selects an optimal set of point correspondences from two
point sequences describing the catheters in two fluoroscopic images. The
selected correspondences are then used for computation of the 3-D re-
construction. The evaluation on 33 clinical biplane images yielded an
average backprojection error of 0.4 mm ± 0.6 mm.

1 Introduction

Radio-frequency catheter ablation can be used to treat atrial fibrillation, cur-
rently being the most common heart arrhythmia. Ablation is usually performed
under fluoroscopic guidance using a C-arm system. Unfortunately, in many cases
physicians can only see the catheters, while the left atrium remains invisible. Of-
ten a 3-D overlay is used to outline the structure of the heart [6]. In contrast to
electro-anatomical mapping systems, C-arm systems provide only 2-D projection
images but no direct 3-D information of the catheters. Physicians need to derive
the respective catheter positions from fluoroscopic imaging mentally. Currently,
point localization from two views is the only way to verify a catheter position.
Electro-anatomical mapping systems provide a 3-D visualization of the catheters
inside a pre-operative data set. To provide the same functionality, we propose a
novel semi-automatic method to reconstruct a catheter used in electrophysiol-
ogy (EP) procedures from two views. In addition, multiple reconstructions of the
mapping catheter at different pulmonary veins (PVs) or at different positions
at one PV can be used for registration. Previously, a method for initial regis-
tration using the catheter in the coronary sinus was proposed [5]. Our proposed

---

* Corresponding author.

method would help to use these methods in clinical practice. Furthermore, the 3-D catheter model could be used as input for other algorithms used for catheter tracking [2] or motion compensation [4]. Previous methods dealed only with guide wires or catheters with low curvature [3,10]. Blob detection can be used to detect the electrodes of the catheter, but not its entirety [8]. Our novel method provides the capability to achieve a 3-D reconstruction of a complex catheter such as the circumferential mapping catheter by requiring only minimal user input. User interaction is limited to a single click on the catheter in each of the imaging frames. This contribution is structured as follows. In the second section, details on the catheter detection are provided. In the third section, the catheter reconstruction is presented. Our evaluation and results are presented in section four. In the final section, we discuss our results and draw some conclusions.

2   Catheter Detection

Catheter detection is performed in three steps: (i) a seedpoint on the transition from the shaft to the catheter tip is marked. Based on image pre-processing, possible parts of the catheter are identified, and (ii) combined to a graph. (iii) a set of edges of the graph is selected. This provides us with a final estimate for the overall catheter. In step (i), a medialness filter [9] is applied to the image. Then a binary image is generated using a dynamic threshold. Before the skeletonization of the binary image is computed, an opening and closing operation is performed to close small gaps. The search space for the catheter is then reduced by selecting the 150,000 pixels from the $1,024 \times 1,024$-image that are nearest to the seedpoint and the skeleton. The skeleton contains the parts of the image which belong most likely to the catheter. In step (ii), the skeleton is transformed into a graph to obtain an analytical representation for a catheter model. The edges of the graph are point sequences that basically follow the skeleton. The nodes of this graph are the points where a point sequence starts or meets another point sequence. An example of such a graph is given in Fig. 1(b). For the computation of this graph, a cost function is defined using the filtered image and a distance transformation of the skeleton. Using this cost function, the shortest paths from all pixels to the seedpoint are computed by Dijkstra's algorithm. The circles and lines of the skeleton are then integrated in the graph. Since the skeleton is not continuous but may contain gaps, the above computed shortest paths from certain feature points to the seedpoint are used as edges for the new graph. The feature points are computed such that the shortest paths from these points to the seedpoint form the circles and lines of the skeleton as well as certain additional connections between parts of the skeleton. For the circles of the skeleton, the two points on the circle which are farthest away from the seedpoint are used as feature points. These points can easily be found by searching for points on the skeleton which have only neighboring skeleton pixels with a lower distance to the seedpoint, see Fig. 1. The shortest paths from these points to the seed point will either run in clockwise or counterclockwise direction, respectively. The remaining lines of the skeleton are inserted into the graph by including the edges belonging
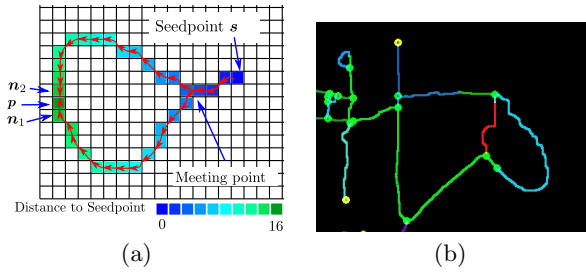
Fig. 1. (a) Circles can be found by considering pixels at which two paths from different directions meet. The color indicates the distance to the seedpoint. The two neighboring pixels of p are closer to the seedpoint than p itself. In such a case, two paths from different directions meet at pixel p. By selecting p and n_2 as feature points, the circle is given by their respective paths to s (b) An example of the generated graph. Each path from a feature point to a node is given in a different color to make them distinguishable.

to the shortest paths from endpoints of the skeleton to the seedpoint. Finally, connections between parts of the skeleton are added to overcome gaps in the skeleton. The connections are found similarly to the circles but this time the points with higher distance to all neighboring points do not need to be part of the skeleton. For each connection, the distance l_b in pixels to the existing graph is computed as well as the distance l_c to the meeting point. Finally, the endpoints of edges in the graph which are neighbored in the image are connected such that circles arise. In step (iii), the above computed graph serves as search space for the catheter. With the seedpoint, some information is given by the physician. The search for the catheter is split up into the search for the catheter shaft and the catheter tip. Due to the standard setups of the C-arm used in electrophysiology procedures, prior knowledge of the image orientation can be taken into account. The shaft of the catheter is considered to enter the image at the bottom edge. Therefore all shortest paths on the graph from the seedpoint to points in the lower half of the image are considered as candidates for the shaft. They are examined with respect to their length, curvature, direction and ending angle. The best shaft candidate maximizes the length while minimizing the curvature. Its direction as well as its ending angle are as vertical as possible. The elliptical catheter tip is determined by considering all possible paths starting from the seedpoint that were not used for the shaft. For selecting the best path, an ellipse is fitted to the path and the ellipse that approximates the elliptical part best is chosen, according to their response to the medialness filter. We also try to make sure that the shape of the path is as elliptical as possible. This is measured by considering the average distance from the ellipse to the path and vice versa. Additionally, the size of the approximated ellipse is used to reject candidates for which the ellipse is not plausible with respect to the known catheter shape. Since the search space for the ellipse search is very large, a population based search strategy is used [1]. In this strategy, an initial population of short paths starting from the seedpoint is evaluated. In each iteration the offspring which continues the paths from the previous generation is evaluated as described above. If a
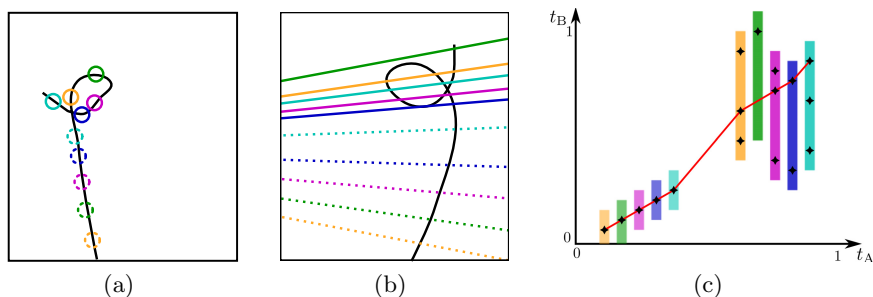
Fig. 2. The spline in (a) is sampled and for each sampling point, possible point correspondences are compute by intersecting the epipolar line with the spline in (b). For each possible point correspondence, the spline parameters are stored (c). A correspondence function that computes for each spline point $C_A(t_A)$ the corresponding spline point $C_B(t_B)$ has to be monotonic. The optimal correspondence function interpolates as much as possible correspondence candidates.

population limit is reached, the paths with the worst fitness value drop out of the population. The search stops, if in the previous three generations no better solution was found. The final result is a 2-D spline along the detected catheter.

3   Catheter Reconstruction

The catheter is reconstructed in 3-D from the two 2-D splines, $C_A(t_A)$ and $C_B(t_B)$, in image plane A and B. They run from bottom to top, i.e. $C_A(0)$ denotes the lowest point at the shaft, $C_A(1)$ is the last point at the tip. Reconstruction is performed by triangulation of corresponding points. To find them, $C_A(t_A)$ is equidistantly sampled and for each point, potential corresponding points are computed by intersecting the epipolar line with the spline $C_B(t_B)$ in image B. For objects with high curvature, a point in A might yield multiple intersections in B. So, a strategy for selecting one of those intersections has to be established. Due to inaccuracies in a practical scenario, the epipolar line might intersect the spline not at the actual corresponding point. As shown by the solid green line in Fig. 2, intersections can, however, occur at other points. A greedy selection strategy as proposed by Baert et al. [2] might yield wrong results in such a case. It selects the intersection point next to the previous one as corresponding point. If one point would be chosen incorrectly, all subsequent points would be incorrect as well, due to the linearity constraint. To avoid such a problem, the selection process is performed after the computation of all correspondence candidates. This is achieved by computing an optimal correspondence function $f : t_A \rightarrow t_B$ that takes all possible point correspondences (t_A, t_B) into account. It should map the first spline parameter to the second and be able to deal with missing points. The monotonic function that contains as many as possible correspondence pairs is selected, see Fig. 2. It might be necessary to skip some points in order to include following points with regard to the monotony constraint. An estimate of missing correspondence points can be computed by interpolating f. The computation
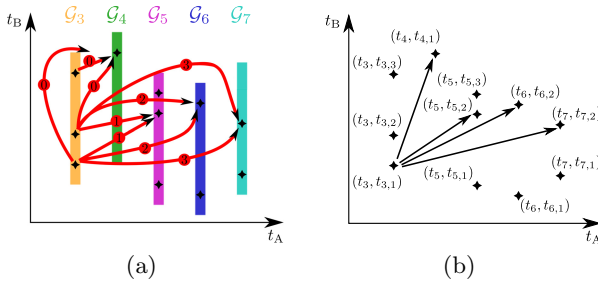
Fig. 3. (a) The optimal correspondence function can be found by computing a shortest path in a graph. The nodes of the graph are the correspondence candidates and the edges connect them such that monotony constraint is preserved. The cost of an edge is equal to the number of skipped points. (b) An example of the graph, the edges are visualized only for node $(t_3, t_{3,1})$.

of this function can be considered as a shortest path problem. A node of the graph corresponds to a possible point correspondence pairs $(t_i, t_{i,j})$. t_i denotes the spline parameter of the i-th sample point of C_A. For a fixed point $C_A(t_i)$, the possible corresponding points are then given by $C_B(t_{i,1}), \dots, C_B(t_{i,m})$, assuming m intersection points with the spline in image plane B. All nodes with the same index i form a group $\mathcal{G}_i$. The edges are designed such that the monotony constraint is preserved. The node $(t_i, t_{i,j})$ is connected by a directed edge to a single node of each group $\mathcal{G}_{i+l}$ with a group jump $l > 0$. The target node of the edge is the node with the smallest index $t_{i+l,k}$ that fulfills the constraint $t_{i+l,k} > t_{i,j}$. Including an edge from group $\mathcal{G}_i$ to a group $\mathcal{G}_{i+l}$, $l \geq 2$ means that the possible point correspondences for the sample points in between are neglected. As only as few points as possible should be omitted, skipping a group is penalized by a cost of 1, so the cost of an edge is $l - 1$.

4   Evaluation and Results

For the evaluation of the detection, clinical data consisting of 33 biplane sequences from 27 patients were used. For each of the resulting 66 monoplane sequences, a detection was computed and compared with a gold standard segmentation. The error was measured as the distance from detected catheter pixels to the gold standard center line. For the generation of the search graph, it turned out to be sufficient to include only those circles in the graph for which the piece from a feature point to the meeting point is longer than 75 pixels. For the connections between parts of the skeleton, taking only those with $l_c > 75$ pixels and $l_c/l_b > 10$ was a good choice. The search for the catheter tip started with a path length of 100 pixel and stopped at latest if a length of 600 pixel was reached. A possible detection result is given in Fig. 4. For the whole catheter, the mean error of the detection is $0.7 \, \text{mm} \pm 2.0 \, \text{mm}$ and the min and max error is $0.0 \, \text{mm}$ and $19.9 \, \text{mm}$. The pixel spacing is $0.1725 \, \text{mm}$, so, the mean error corresponds to 4.1 pixels in the 2-D images. For the tip only, the mean error is $0.4 \, \text{mm} \pm 0.3 \, \text{mm}$,

(a)   (b)
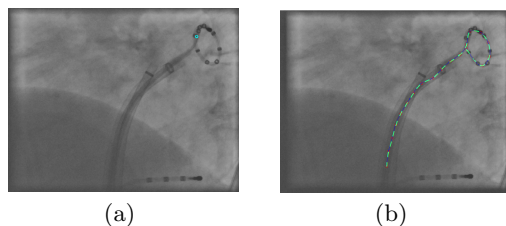
Fig. 4. (a) Manual initialization of the catheter detection. The seedpoint of the proposed method is set by the physician at the transition between the shaft and the catheter tip. (b) Resulting spline after detection of the catheter.

(a)   (b)
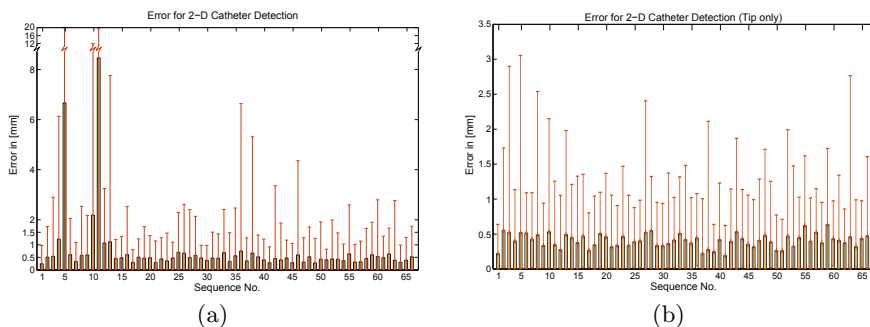
Fig. 5. Mean detection errors for each sequence, the thin bars denote the maximal and minimal error. In (a) the detection error for the whole catheter is given. The total mean error is 0.7 mm ± 2.0 mm. (b) shows the error for the catheter tip. The total mean error is 0.4 mm ± 0.3 mm.

the min and max error is 0.0 mm and 3.1 mm, see Fig. 5. As the catheter tip is the important part for the physician, our results meet current clinical requirements. The average percentage of undetected pixels was 6.0 % ± 10.2 %. The min and max percentage of undetected pixels were 0.0 % and 44 %, respectively. The average percentage of false positive pixels was 1.3 % ± 7.1 % The min and max percentage of false positives were 0.0 % and 47 %, respectively. The evaluation of the reconstruction was performed using clinical data. Since no clinical 3-D data was available as ground truth, the result of the 3-D reconstruction was backprojected into the image planes and the error of the reprojected points to a gold standard segmentation was computed. The resulting mean error was 0.4 mm ± 0.6 mm and 2.3 pixels, respectively, for the whole catheter, and 0.7 mm ± 0.7 mm for the catheter tip. The min and max error was 0.0 mm and 4.8 mm in both cases. Using the results of the detection for reconstruction, the result differed to the reconstruction using the gold-standard by 1.5 mm ± 2.5 mm, the min and max difference was 0.0 mm and 15.7 mm. Fig. 6 shows the results for each biplane pair. The detection requires on average 8.7 sec and the reconstruction 0.15 sec on an Intel i7 with 2,67 GHz and 4 GB RAM. The medialness filter computation was performed on a NVIDIA Quadro FX 880M GPU.
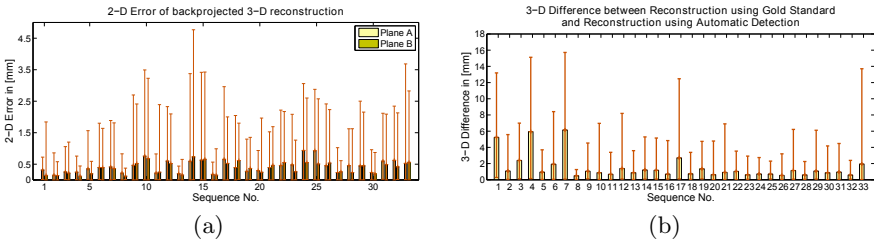
Fig. 6. (a) shows the mean error of the reprojected 3-D reconstruction for each sequence. The overall mean error is 0.4 mm ± 0.6 mm. (b) shows the deviation of the reconstruction using the detection results to the reconstruction using the gold standard segmentation. The mean deviation is 1.5 mm ± 2.5 mm.

Fig. 7. Reconstruction result using the previous approach (a) and our new method (b). While both methods perform well for the linear part of the catheter, the previous approach fails for the curved catheter tip.

5   Discussion and Conclusions

The results of the catheter detection show that detection of the whole catheter is possible if only a single seedpoint is provided. The mean error is 0.7 mm and therefore sub-millimeter accuracy can be reached for catheter parts that were detected by the algorithm, thus, enabling catheter localization for EP applications [7]. The percentage of undetected catheter parts is with 6.0 % lower than the methods proposed by Honnorath et al. [10] and Franken et al. [8]. Also the number of false positive pixels is lower when compared to [10]. Compared to the previous approach by Baert [2], our new approach can deal better with curved catheters as shown in Fig. 7. In some cases, two catheters may be overlapping and the wrong catheter is selected as shaft. In such cases, the physician could manually adjust the detection result before performing the 3-D reconstruction.

References

1. Bäck, T., Hammel, U., Schwefel, H.: Evolutionary computation: Comments on the history and current state. IEEE Tr. on Evolutionary Comp. 1(1), 3–17 (1997)
2. Baert, S., van de Kraats, E., van Walsum, T., Viergever, M., Niessen, W.: Three-dimensional guide-wire reconstruction from biplane image sequences for integrated display in 3-D vasculature. IEEE Tr. on Medical Imaging 22(10), 1252–1258 (2003)
3. Bender, H.-J., Männer, R., Poliwoda, C., Roth, S., Walz, M.: Reconstruction of 3D catheter paths from 2D X-ray projections. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 981–989. Springer, Heidelberg (1999)
4. Brost, A., Liao, R., Strobel, N., Hornegger, J.: Respiratory motion compensation by model-based catheter tracking during EP procedures. Medical Image Analysis 14(5), 695–706 (2010)
5. Brost, A., Bourier, F., Yatziv, L., Koch, M., Hornegger, J., Strobel, N., Kurzidim, K.: First steps towards initial registration for electrophysiology procedures. In: Wong, K.H., Holmes III, D.R. (eds.) Medical Imaging 2011: Visualization, Image-Guided Procedures, and Modeling, SPIE, vol. 7964, p. 79641 (2011)
6. Dilling-Boer, D., van der Merwe, N., Adams, J., Foulon, S., Goethals, H., Willems, R., Ector, H., Heidbüchel, H.: Ablation of Focally Induced Atrial Fibrillation. Journal of Cardiovascular Electrophysiology 15(2), 200–205 (2004)
7. Esteghamatian, M., Azimifar, Z., Radau, P., Wright, G.: Real-time 2D-3D MR cardiac image registration during respiration using extended Kalman filter predictors. In: Baozong, Y. (ed.) 9th International Conference on Signal Processing, ICSP 2008, pp. 1325–1328. IEEE (2008)
8. Franken, E., Rongen, P., van Almsick, M., ter Haar Romeny, B.: Detection of Electrophysiology Catheters in Noisy Fluoroscopy Images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006, Part II. LNCS, vol. 4191, pp. 25–32. Springer, Heidelberg (2006)
9. Gülsün, M., Tek, H.: Robust Vessel Tree Modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 602–611. Springer, Heidelberg (2008)
10. Honnorat, N., Vaillant, R., Paragios, N.: Guide-Wire Extraction through Perceptual Organization of Local Segments in Fluoroscopic Images. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 440–448. Springer, Heidelberg (2010)

Feature Classification
for Tracking Articulated Surgical Tools

Austin Reiter[1], Peter K. Allen[1], and Tao Zhao[2]

[1] Columbia University, New York NY USA
{areiter,allen}@cs.columbia.edu
[2] Intuitive Surgical, Inc, Sunnyvale CA USA
tao.zhao@intusurg.com

Abstract. Tool tracking is an accepted capability for computer-aided surgical intervention which has numerous applications, both in robotic and manual minimally-invasive procedures. In this paper, we describe a tracking system which learns visual feature descriptors as class-specific landmarks on an articulated tool. The features are localized in 3D using stereo vision and are fused with the robot kinematics to track all of the joints of the dexterous manipulator. Experiments are performed using previously-collected porcine data from a surgical robot.

Keywords: Tool Tracking, Surgical Robotics, Learning, Features, Fusion.

1 Introduction

Robotic surgery has become widely used in recent years as it has enhanced the inherent abilities of the human surgeon. For example, there are more than 1800 da Vinci® [1] surgical systems working in operating rooms all over the world that performed about 360,000 surgical procedures in 2011. High definition stereo vision helps the surgeon see the anatomy and interact with the surgical tools with great clarity. Augmenting the surgeon's vision with other relevant information in the form of graphical overlays can further help the surgeons/patients in a different dimension.

Knowledge of the locations of tools in the endoscopic video can enable a wide spectrum of applications. Virtual measurements can be utilized to provide accurate measurement of sizes of various anatomical structures. Virtual overlays indicative of the status of the tool (e.g., the firing status of an electro-cautery tool) can be placed at the tip of the instrument which is close to the surgeon's visual center of attention, enhancing the safety of using such tools. It is also useful in managing the tools that are off the screen, increasing patient's safety, or for visual servoing of motorized cameras.

The joints of a robotic surgical system are usually equipped with encoders so that the pose of the end effectors can be computed using forward kinematics. On one hand, the kinematics chain between the camera and the tool tip involves 18 joints and more than 2 meters in cumulative length, which is challenging to the accuracy of absolute position sensing. On the other hand, a master-slave system does not require high absolute accuracy because surgeons are in the control loop. As a result, we have observed up to 1-inch of absolute error, which is too large for most of the applications that are

mentioned above. Therefore, detecting and tracking the tools from images is a practical way to achieve the accuracy requirements of the applications.

Previous approaches to tool tracking have employed specialized fiducial markers to locate the tool [2,3,4]. Effective as they are, there are practical challenges such as manufacturability and cost. Alternatively, marker-less techniques have used: color segmentation to label pixels in the image as tool against the background [5,6,7]; geometric priors to confine the search space from the abdominal wall [8,9,10]; or combining different features together to detect the tool in the image [11,12,13]. Most prior work emphasizes the shaft, however in robotic surgery surgeons tend to work very close to the anatomy and the small visible part of the shaft may cause these algorithms to work poorly. This has motivated us to investigate the features on the tip of the tools.

In this paper we present a tracking system which learns the appearances of natural landmarks on an articulated tool by training an efficient multi-class classifier on a discriminative feature descriptor. We run the classifier on an image frame to detect all extrema representing the location of each feature type, where confidence values help reject false positives. We stereo match in the corresponding camera to recover 3D locations. By also knowing these landmark locations on the tool CAD model, we recover a pose offset of the kinematics using an Extended Kalman Filter. This fusion of vision and kinematics fills in the gap of missed vision detections and the articulations that are not estimated by vision, making the estimation of the tool pose available at all times.

2 Materials and Methods

2.1 Training Data Collection

We begin by collecting data to train our classifier. We use five different video sequences which span various in-vivo experiments, to best cover a range of appearance and lighting scenarios. We concentrate on the Large Needle Driver (LND) tool, keeping in mind this technique may be applied to any other types of tools. Seven visually salient and stable landmarks are manually selected and shown on the left of Fig. 1. The features chosen are of the pins that hold the distal clevis together, the **IS** logo in the center, and the wheel and wheel pin. We also know these locations on the tool's CAD model, which will be used for association to compute the final articulated pose.

For each frame in the labeling procedure, we manually drag a bounding-box around each feature of interest, being careful to avoid contamination from pixels which don't belong to the tool. Overall, we use $\sim 15,000$ training samples across the 7 classes.

2.2 Feature Descriptor

There has been a large amount of prior work on visual feature detection and matching in the computer vision community. Scale and affine invariant feature descriptors [15] have been very successful in matching planar features. However, we found that they work poorly for features on metal surfaces with lighting changes, as in the case of surgical tools with varying poses and light directions. The right-side of Fig. 1 shows example appearance changes typically encountered of the *IS_Logo* feature. We require
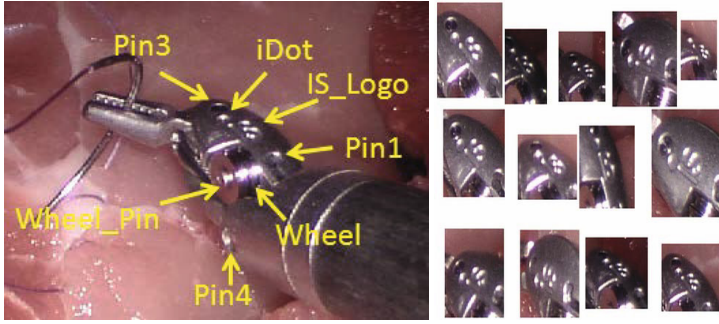
Fig. 1. [**Left**] The feature classes we detect on the LND tool. We concentrate on 7 different types of naturally-occurring landmarks. [**Right**] Examples appearance changes of the *IS_Logo* feature.

a discriminative and robust descriptor for our feature classes as each is fairly small (e.g., 17-25 pixels wide), and so we chose the Region Covariance Descriptor [16], where the covariance matrix of d features in a small image region serves as the feature descriptor. Given an image I, we extract $d=11$ features, resulting in the feature image **F**:

$$\mathbf{F} = [x \; y \; \mathbb{H} \; \mathbb{S} \; \mathbb{L} \; I_x \; I_y \; I_{xx} \; I_{yy} \; \sqrt{I_x^2 + I_y^2} \; \arctan(I_y/I_x)] \tag{1}$$

where x, y are the pixel locations; $\mathbb{H}, \mathbb{S}, \mathbb{L}$ are the hue, saturation, and luminance values at pixel location (x, y); I_x, I_y are the 1st-order spatial derivatives; I_{xx}, I_{yy} are the 2nd-order spatial derivatives; and the latter two features are the gradient magnitude and orientation, respectively. The covariance matrix $C_\mathbf{R} \in \mathbb{R}^{d \times d}$ of an arbitrary rectangular region **R** within F then becomes our feature descriptor.

Each $C_\mathbf{R}$ can be computed efficiently using integral images. We compute the sum of each feature dimension as well as the sum of the multiplication of every two feature dimensions. Given these first and second order integral image tensors, it can be shown that the covariance matrix of *any rectangular region* can be extracted in $O(d^2)$ time [16]. Using the ground truth data from Sec. 2.1, we extract covariance descriptors of each feature and store the associated feature label for training a classifier.

2.3 Feature Classification

There are several multi-class classifiers which may suit this problem. We adapt a method called *Randomized Trees* (RTs) [17] due to its computational efficiency. In addition to providing feature labels, we would like to retrieve confidence values for the classification task which we use to construct class-conditional likelihood images for each class.

RTs naturally handle multi-class problems very efficiently while retaining an easy training procedure. The RT classifier Λ is made up of a series of L randomly-generated trees $\Lambda = [\gamma_1, \ldots, \gamma_L]$, each of depth m. Each tree γ_i, for $i \in 1, \ldots, L$, is a fully-balanced binary tree made up of internal nodes, each of which contains a randomly-generated test that splits the space of data to be classified, and leaf nodes which contain estimates of the posterior distributions of the feature classes.

To train, the training features are dropped down the tree, performing binary tests at each internal node until a leaf node is reached. Each leaf node contains a histogram of length equal to the number of feature classes b. The histogram at each leaf counts the number of times a feature with each class label reaches that node. At the end of the training session, the histograms are normalized into probabilities using the total number of hits at that node. A feature is then classified by dropping it down the trained tree, again until a leaf node is reached, and is assigned the probabilities of belonging to a feature class depending on the posterior distribution stored at the leaf from training.

Because it's computationally infeasible to perform all possible tests of the feature, L and m should be chosen so as to cover the search space sufficiently and avoid randomness. Although this approach has been very successful for matching keypoints [17], traditionally the internal node tests are performed on a small patch of the gray image by randomly selecting 2 pixel locations and applying a binary operation ($\leq$) to determine which path to take to a child. In our problem, we are using feature descriptor vectors rather than image patches, and so we must adapt the node tests to suit our problem.

To this end, we use a similar approach to [18] for the node tests. For each internal tree node we construct a random linear classifier $h_i(\mathbf{x})$ on feature vector $\mathbf{x}$ to split:

$$h_i(\mathbf{x}) = \begin{cases} \mathbf{n^T x} + z \leq 0 & \text{go to right child} \\ otherwise & \text{go to left child} \end{cases} \quad (2)$$

where $\mathbf{n}$ is a randomly generated vector of the same length as feature $\mathbf{x}$ with values in the range $[-1, 1]$ and $z \in [-1, 1]$ is also random. We found that this node test allows for robust splitting of the data and is computationally efficient. In this way, we build up probability distributions at the leaf nodes with the training descriptors. The results from each γ_i are averaged across the L trees. However, the d-dimensional nonsingular covariance matrices cannot be used *as is* to perform this task directly because they do not lie on a vector space, but rather on a connected Riemannian manifold. Due to space limitations, we refer the reader to [19] for a mathematical overview on post-processing the covariance descriptors to a Euclidean vector-space for use with a classifier. In the end, our $[d \times d]$ dimensional matrices $\mathbf{C_R}$ are mapped to vectors $\mathbf{c}_j \in \mathbb{R}^{d(d+1)/2}$.

2.4   Feature Class Labeling and Reconstruction

Labeling. Given our trained classifier Λ, we detect features for each label in an image by computing covariances $\mathbf{C_R}$, each of which is mapped to a vector space, producing $\mathbf{c}_j$. We drop each $\mathbf{c}_j$ through the trees γ_i and average the probabilities at the obtained leaf nodes to get a probability distribution p_L, representing the probability of $\mathbf{c}_j$ belonging to each of the L feature classes, resulting in L class likelihood images. To get the pixel locations, we perform non-maximal suppression in each class likelihood image. Example detections and likelihoods are shown in Figs. 2(a) and 2(c)-2(i), respectively.

Although the integral images afford efficient extractions of the covariances, we can reduce the computations further by initially segmenting the pixels of the image to identify areas of interest to classify. Using the method in [11], we train a Gaussian Mixture Model of several color and texture features to classify pixels into 1 of 3 groups (metal,

| (a) Class Extrema | (b) Mask Prior | (c) iDot | (d) IS_Logo | (e) Pin1 |

| (f) Pin3 | (g) Pin4 | (h) Wheel | (i) Wheel_Pin |
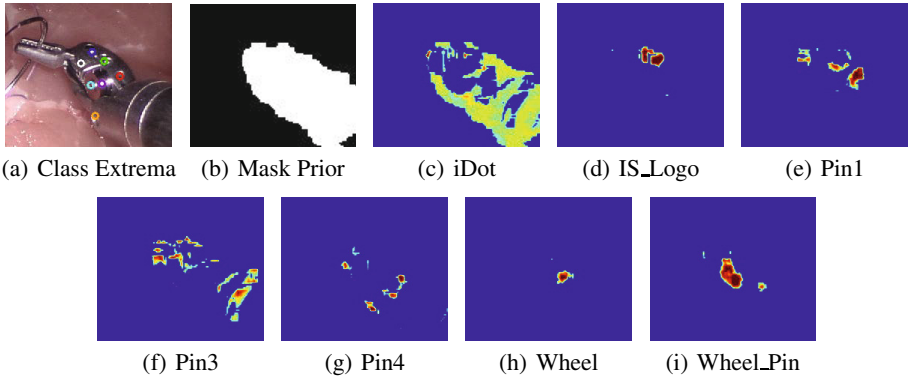
Fig. 2. Example likelihood images along with 6/7 successfully detected feature classes correctly located as extrema in (a). The *Pin3* (f) feature is incorrectly localized (white circle). The color-coding for the circles in (a) is: **Blue** for iDot, **Green** for IS_Logo (d), **Red** for Pin1 (e), **Orange** for Pin4 (g), **Purple** for Wheel (h), and **Cyan** for Wheel_Pin (i). A mask prior is shown in (b) which detects pixels on the metal part of the tool to reduce the number of pixels needing classification.

shaft, and background) and use the metal likelihood to produce a binary image of pixels to run through our classifier Λ. Fig. 2(b) shows a mask prior for the video image in 2(a).

3D Reconstruction. Now that we have candidate pixel locations for each feature class, we retrieve the 3D locations by stereo matching the feature tracks in the corresponding stereo camera using normalized cross-correlation checks along the epipolar line and triangulating the features. These feed into our fusion stage, described next.

2.5   Vision and Kinematics Fusion

The robot kinematics data can be fused with the vision to fill in the gap of the missed detections and to facilitate rejection of outliers. Many surgical robots like the da Vinci® need to maintain a stationary insertion point (also termed *remote center of motion*, or RCM). The errors in the passive setup joints accumulate at the RCM, resulting in a pose offset between the actual pose of the RCM and the pose computed using forward kinematics. This pose offset should be constant or slowly changing for a given surgical setup. Therefore, we can fuse vision and kinematics by solving this 6-DOF pose offset.

We denote **RCS** as the coordinate system of the true RCM and **KCS** as the coordinate system of the RCM according to the kinematics. The coordinates of a point **p** in both coordinate systems are associated by $\mathbf{p}^K = \mathbf{R}_R^K \mathbf{p}^R + \mathbf{c}_R^K$. We use an Extended Kalman Filter (EKF) to estimate the rigid transformation $\mathbf{R}_R^K, \mathbf{c}_R^K$ due to its adaptive nature, computational efficiency and the ability for uncertainty propagation. The information form of the filter is chosen to easily deal with a varying number of detected features. Also, the filter does not require the solution to be fully determined in a single frame. We omit the details of the implementation due to space limitations and refer the interested reader to [20] for more details.

It is possible that the output of the feature classification contains outliers. We employ RANSAC to enforce the rigid transformation using a sliding window approach and so
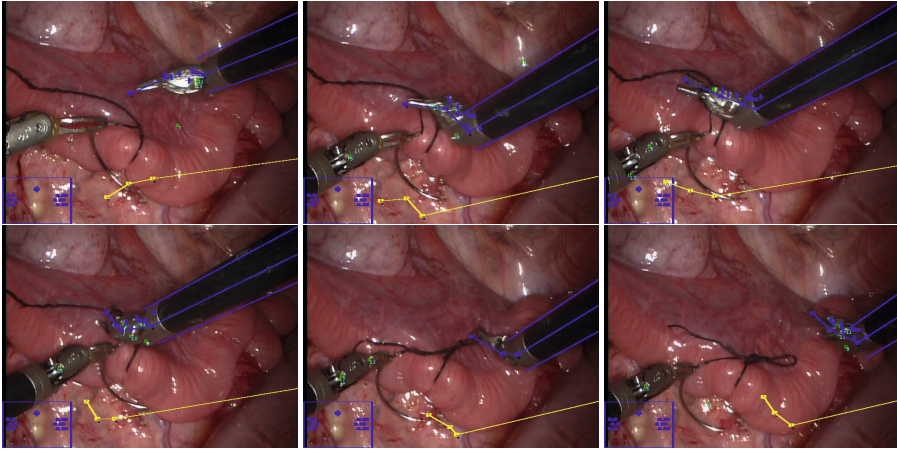
Fig. 3. Examples from our fusion tracker with the overlays for both the raw (yellow) and corrected (blue) kinematics. A visual inspection shows the tracker fixing the kinematics quite accurately.

only the inliers are fed into the EKF. We require a minimum of $\sim 30 - 50\%$ inliers to begin the filtering procedure.

3 Results

We experimented on previously collected porcine data from a da Vinci® surgical robot. The data which was used to test was specifically not included in the training procedure. After the off-line training of our RT classifier Λ (Sec: 2.3) using the seven feature classes shown in Fig. 1, we detected features (Sec. 2.4) and performed 3D reconstruction. These point locations were fed into the fusion module (Sec. 2.5) and the final kinematics joint overlay is drawn in the original image frame.

Six sample results are shown in Fig. 3, in which the yellow lines represent the raw kinematics projected into the image frames and the blue lines show the *fixed* kinematics resulting from our tracker. Notice the significant errors in some cases, motivating the need for the ideas presented in this paper. A visual inspection yields a fairly accurate correction of the kinematics overlaid on the right-most tool. The bottom-row middle-column shows a case where no features are visible, yet the EFK can predict because the remote center bias was previously accounted for correctly and remained static. This affords longer overall tracking times. On average it required $\sim 10 - 12$ frames for the fusion module to collect enough evidence to lock on to the true instrument kinematics.

Accuracy. The accuracy requirements for tool tracking depends on the application. Some applications, such as virtual measurement, may require ≤ 1mm accuracy while others such as status overlay can be more tolerant. It is regarded as acceptable if the estimated tool centerline is within the tool shaft, so it is not confused with another tool. This has motivated our evaluation method shown in Fig. 4, without being able to
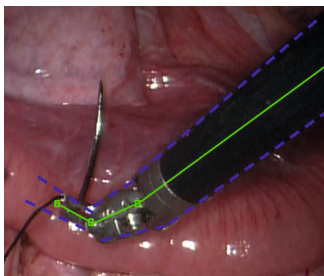
Fig. 4. To evaluate our kinematics estimates, the projected overlays must fall within the boundaries labeled as dotted blue lines here.

obtain ground truth in-vivo. The dotted blue lines define an acceptable boundary for the camera-projection of the kinematics. We manually inspect each frame of the test sequences, and resulted in a 93% accuracy rate over 1600 test frames. We attribute the incorrect estimates to the initial time necessary to lock on to the tool while the fusion module gathers evidence and occasional poor localization of the 2D features.

Occasionally we notice that the estimate is accurate on the tool tip, but slightly offset on the shaft. This occurs because we only include one feature on the shaft (*Pin4*), but in the future we will look to include more shaft information. Concurrent with this work we performed a study [21] on the feature detection accuracy, where we obtained an average localization accuracy of 86%, although this varies depending on the feature type. We also note that although some feature types are not always detected, we need only $\sim 3 - 4$ on a given frame because of the fusion, and so across the 7 chosen landmarks our experiments show that the percent correct achieved is sufficient for long-term tracking.

Timing. The tracker runs at ~ 1.2 secs/frame, where the feature detection takes most of the processing. This is dependent on: number of trees in Λ (90) , depth of each tree γ_i (10), number of features used in $\mathbf{C_R}$ (11), and the quality of the initial segmentation. It is also worth mentioning that we are estimating a pose offset that changes very slowly, therefore it is not critical that the estimation runs at frame-rate. Future speed-ups can come from GPU enhancements, because many parts of the algorithm are inherently parallel, and *tracking* of the features (between consecutive frames, the features shouldn't move far and so we only need to classify small windows around previous detections).

4   Conclusion

In this paper we have presented a method to learn to detect naturally-occurring landmarks on an articulated surgical tool to track it over time. We showed robustness in both feature detection and fusion of the vision observations with the inaccurate robot kinematics. Future work includes testing on different types of tools as well as tracking multiple tools in the scene simultaneously.

References

1. Intuitive Surgical, Inc., http://www.intuitivesurgical.com/
2. Wei, G.-Q., Arbter, K., Hirzinger, G.: Automatic Tracking of Laparoscopic Instruments by Color Coding. In: 1st Joint Conf. on Computer Vision, Virtual Reality, and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery, pp. 357–366 (1997)
3. Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M., Morel, G., Leroy, J., Soler, L., Marescaux, J.: Autonomous 3-D Positioning of Surgical Instruments in Robotized Laparoscopic Surgery Using Visual Servoing. IEEE Trans. on Robotics and Automation 19(5), 842–853 (2003)
4. Groeger, M., Arbter, K., Hirzinger, G.: Motion Tracking for Minimally Invasive Robotic Surgery, pp. 117–148. Medical Robotics ITech Education and Publishing (2008)
5. Lee, C., Uecker, D.: Image Analysis for Automated Tracking in Robot-Assisted Endoscopic Surgery. In: Intl. Conf. on Computer Vision and Image Processing (1994)
6. Doignon, C., Nageotte, F., de Mathelin, M.: Detection of grey regions in color images: application to the segmentation of a surgical instrument in robotized laparoscopy. In: IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (2004)
7. Doignon, C., Graebling, P., de Mathelin, M.: Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature. Real-Time Imaging (2005)
8. Doignon, C., Nageotte, F., de Mathelin, M.: Segmentation and Guidance of Multiple Rigid Objects for Intra-operative Endoscopic Vision. In: Vidal, R., Heyden, A., Ma, Y. (eds.) WDV 2005/2006. LNCS, vol. 4358, pp. 314–327. Springer, Heidelberg (2007)
9. Voros, S., Long, J.-A., Cinquin, P.: Automatic Detection of Instruments in Laparoscopic Images: A First Step Towards High-level Command of Robotic Endoscopic Holders. Intl. J. of Robotics Research 26(11-12), 1173–1190 (2007)
10. Wolf, R., Duchateau, J., Cinquin, P., Voros, S.: 3D Tracking of Laparoscopic Instruments Using Statistical and Geometric Modeling. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 203–210. Springer, Heidelberg (2011)
11. Pezzementi, Z., Voros, S., Hager, G.: Articulated Object Tracking By Rendering Consistent Appearance Parts. In: IEEE Intl. Conf. on Robotics and Automation (2009)
12. Reiter, A., Allen, P.K.: An Online Approach To In-Vivo Tracking Using Synergistic Features. In: IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (2010)
13. Sznitman, R., Basu, A., Richa, R., Handa, J., Gehlbach, P., Taylor, R.H., Jedynak, B., Hager, G.D.: Unified Detection and Tracking in Retinal Microsurgery. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 1–8. Springer, Heidelberg (2011)
14. Burschka, D., Corso, J., Dewan, M., Lau, W., Li, M., Lin, H., Marayong, P., Ramey, N., Hager, G., Hoffman, B., Larkin, D., Hasser, C.: Navigating Inner-Space: 3-D Assistance For Minimally Invasive Surgery. In: IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (2004)
15. Lowe, D.: Distinctive image features from scale-invariant keypoints. Intl. J. of Computer Vision 60(2), 91–110 (2004)
16. Tuzel, O., Porikli, F., Meer, P.: Region Covariance: A Fast Descriptor for Detection and Classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
17. Lepetit, V., Fua, P.: Keypoint Recognition Using Randomized Trees. Trans. on Pattern Analysis and Machine Intelligence 28(9), 1465–1479 (2006)
18. Bosch, A., Zisserman, A., Muoz, X.: Image Classification using Random Forests and Ferns. In: IEEE Intl. Conf. on Computer Vision, pp. 1–8 (2007)

19. Tuzel, O., Porikli, F., Meer, P.: Human Detection via Classification on Riemannian Manifolds. In: IEEE. Conf. on Computer Vision and Pattern Recognition (2007)
20. Zhao, T., Zhao, W., Hoffman, B.D., Nowlin, W.C., Hui, H.: Efficient Vision and Kinematic Data Fusion for Robotic Surgical Instruments and Other Applications. US Patent US 20100331855 (June 2009)
21. Reiter, A., Allen, P.K., Zhao, T.: Learning Features on Robotic Surgical Tools. In: Workshop on Medical Computer Vision, CVPR 2012 (2012)

Image-Based Tracking of the Teeth
for Orthodontic Augmented Reality

André Aichert[1,2], Wolfgang Wein[1,2], Alexander Ladikos[1],
Tobias Reichl[2], and Nassir Navab[2]

[1] White Lion Technologies AG, Munich, Germany
[2] Chair for Computer Aided Medical Procedures (CAMP),
Technische Universität München, Munich, Germany
`aichert@cs.tum.edu`

Abstract. We present image-based methods for tracking teeth in a
video image with respect to a CT scan of the jaw, in order to enable
a novel light-weight augmented reality (AR) system in orthodontistry.
Its purpose is guided bracket placement in orthodontic correction. In
this context, our goal is to determine the position of the patient maxilla
and mandible in a video image solely based on a CT scan. This is suit-
able for image guidance through an overlay of the video image with the
planned position of brackets in a monocular AR system. Our tracking al-
gorithm addresses the contradicting requirements of robustness, accuracy
and performance in two problem-specific formulations. First, we exploit
a distance-based modulation of two iso-surfaces from the CT image to
approximate the appearance of the gum line. Second, back-projection of
previous video frames to an iso-surface is used to account for recently
placed brackets. In combination, this novel algorithm allowed us to track
several sequences of three patient videos of real procedures, despite dif-
ficult lighting conditions. Paired with a systematic evaluation, we were
able to show practical feasibility of such a system.

1 Introduction

This paper suggests a novel solution for guidance in orthodontic applications
with a light-weight monocular video see-through Augmented Reality (AR) sys-
tem. It targets the guided placement of brackets onto individual teeth in order to
improve efficacy and reduce chair time of bracket placement and re-adjustments
for dental braces in orthodontic correction, therefore allowing to incorporate
pre-procedure simulation and planning. The state of the art for this procedure
relies solely on the experience of the orthodontist for both placement of the
brackets and the choice of the wire tension between the brackets. In dentistry,
low-dose cone-beam CT reconstructions of the jaw are typically obtained with
modern digital volume tomography (DVT) devices, with an acceptable dose limit
even for teenagers. Related research [1] has developed simulations based on fi-
nite element methods from such CT data of teeth and bone. Those simulations
could be used in a pre-procedural planning of the optimal bracket placement and

wire tension, such that patient teeth move in an ideal manner while minimizing rotation. While much less accuracy is needed than for example in implant placement, a realization of the pre-procedure plan requires guided placement of the brackets. Augmentation of a patient video showing a superimposition of the newly placed bracket with its planned position would suffice. Potential benefits include higher efficacy due to a reduction in chair time with fewer follow-up visits for corrections, as well as reduced probability of relapse. Therefore, this routine procedure could be improved with a light-weight monocular augmented reality system, while avoiding the cost and complexity of a full-scale medical AR solution.

2   Related Work

One of the central choices which had to be made in realizing this system, is the choice of tracking algorithm [7]. There are several approaches commonly employed for optical tracking, such as marker-based tracking, template-based tracking [2], feature-based tracking [8] and edge-based tracking [5] as well as combinations of these methods. However, due to the nature of our tracking target not all methods can be used. Marker-based tracking is not relevant, because we do not want to augment the scene. When using external tracking systems, a disadvantage besides their high cost is the challenge to keep overall tracking error low, particularly in scenarios such as dental implant placement [12]. This is because the overall system accuracy is limited by the accumulated errors from the tracking system itself, patient registration, hand-eye calibration, synchronization, etc. In order to avoid such an accumulation of errors (as well as additional, expensive equipment) we employ solely image-based tracking methods, which track the patient jaw directly in the video used for the overlay. Template-based tracking by itself is too unstable due to illumination variations, occlusions and the need for an initially textured model of the scene, which we don't have. Feature-based tracking is also not feasible since our scene is mostly textureless and feature-point extractors will not find enough reliable features. This only leaves edge-based tracking methods for use in our system. As can be seen in Fig. 1 edges are a very dominant and a stable feature in the input image. We therefore chose to use edges as our primary tracking modality which is augmented by template-based tracking methods for increased robustness.

3   Methods

In our chosen scenario, a DVT volume of the jaw is always available, since it is the basis of the numerical simulation for planning. Therefore, we investigate methods to automatically align a DVT volume with a video image feed, which relates our problem to medical 2D-3D registration [9] and tracking [11]. For an impression of the scenario see figure Fig. 1. We attempt to use all information available from these two modalities and we present a method consisting of two complementary steps. One step provides a good overall alignment, the other step ensures robust
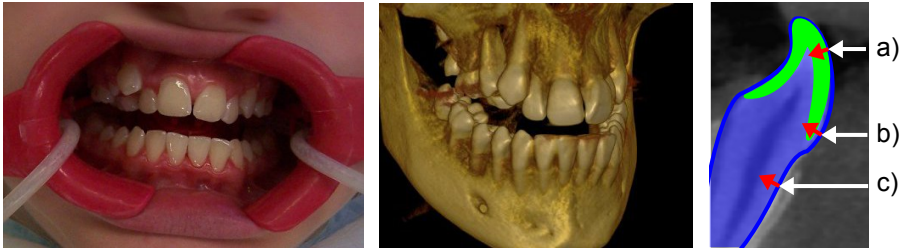
Fig. 1. Photograph of a prepared patient (left), a high-quality volume rendering of the jaw from a DVT volume (center) and schematic of distance-based modulation of two iso-surfaces (right)

tracking even in light of changing conditions. The underlying 2D-3D registration problem is solved through an iterative optimization of a similarity metric over a 6 DOF pose.

3.1   Dual Iso-Surfaces

In order to compute image similarity, we need a fast method of generating a 2D image from the DVT volume for comparison with the 2D video image. While direct volume rendering is able to create close to photo-realistic images, overly complex methods are too slow for real-time registration. Simpler methods, such as non-polygonal iso-surfaces can be computed extremely efficiently, since they represent only one intensity threshold in the data set. In the patient videos, the shape of the teeth and the gum line contain the most reliable geometric information. Unfortunately, in the DVT volume the gum line itself is an interface between two intensities, i.e. enamel and gum. It can therefore not be retrieved by single iso-surface rendering. We suggest using a modulation based on normal distances of two iso-surfaces for the visualization of the gum line. Related work in "Focus and Context" visualization [6] addressed a similar problem for context modulation, where one iso-surface is shown transparent when close to a second one.

While the teeth are easily visualized as the highest CT intensities (i.e. X-Ray attenuation or Hounsfield units), different types of tissue, including the tongue, cheeks, lip and gum all have almost identical attenuation values. Thus, the lip folding over the gum cannot be separated, and therefore the actual gum line cannot be visualized reliably. As an alternative solution, we visualize the interface between enamel and bone or dentin (i.e. the roots of the teeth). This line approximately follows the course of the gum line and we could verify in experiments that this approximation is bias-free with respect to the registration. Please see Fig. 3 (right) for a direct comparison of the dual iso-surface and the textured iso-surface, in particular the course of the gum line.

We efficiently implement this in a single pass of a GPU ray-caster, with a speed close to single iso-surface rendering. As a ray from the camera center into the scene hits the outer surface representing gum (or dentin instead), its direction is changed to the normal of that surface. The ray then is followed for

only a few millimeters more, possibly hitting the second surface for enamel. See Fig. 1 for three example points on the outer surface (blue) and their relation to the inner surface for enamel (green) in a cross section. If the enamel surface is not found (case (c) in Fig. 1 (right)), the gum surface is fully opaque. If the enamel surface is right next to the outer surface (Fig. 1 (right, a)), the outer surface is fully transparent. Otherwise, the surfaces are blended based on their distance (Fig. 1 (right, b)).

3.2   Dissimilarity Metric

Once the DVT volume is rendered, we compare it to the video image. We apply an edge-based similarity metric between the video image and the dual iso-surface rendering. This is based on weighting the distance to the closest edge in the video image with the gradient magnitude of the dual iso-surface rendering. Lower results of this metric indicate that the edges are well-aligned. For efficiency, we compute a distance map from the output of a Canny edge filter [3] of the video frame, since the video image stays constant during the pose optimization. The measure then becomes

$$d = \frac{\sum\limits_{x,y} (d(x,y) \cdot g(x,y))}{\sum\limits_{x,y} g(x,y)}$$

where $d(x,y)$ is the distance map of the video edges and $g(x,y) = \|\nabla_{(x,y)} J\|^2$ is the squared image gradient magnitude of the dual iso-surface rendering J at pixel coordinates (x,y). We use an exponent of two, since we put more emphasis on regions with large gradients (i.e. edges), while reducing the weight of areas with only small variations in gradient caused by noise. Fig. 2 shows a plot of the dissimilarity metric against x and y translations parallel to the image plane. Millimeters

Fig. 2. Plot of the dissimilarity metric d for translations parallel to the image plane with a clear minimum for perfect visual alignment at the center

are measured at the depth of the jaw. Notice a clear minimum at the center of the plot, which represents perfect visual alignment such as seen in Figure 3. Dissimilarity increases more quickly in y-direction, since the dominant occlusion edges of the teeth are in x-direction in the image. We use an elliptical region of interest, defined by two focal points as the projections of the canines.

3.3   Initial Alignment

In order to increase capture range and speed up the process, we propose another step complementing this edge-based approach. Note that there is a variety of

Fig. 3. Video image overlaid with an aligned iso-surface (left) and examples of the rendering methods suggested in Sections 3.3 and 3.1 (center) and a flowchart of the prototype system (right)

established tracking techniques that can be used to achieve such a pre-alignment, some may not even need knowledge of the CT.

Since in our scenario the CT is available, we can reproject video pixels into arbitrary views based on the surface shape from the CT and correct alignment on any one video image. If we choose a texture wrapping of a 2D texture for the iso-surface, we can also project this color information directly to the surface. With a textured model of the jaw, we obtain a much simpler mono-modal registration problem, which can be solved using a simple mono-modal similarity metric. We use a simple frontal linear projection of the texture image to the surface. This is acceptable for this application, since both patient motion and changes in view direction are small, as the orthodontist remains on the same side of the patient.

The system can be initialized manually, by asking the orthodontist to move their head and roughly align a transparent view of the CT with his own vision. The system uses edge based registration to get an initial alignment, which in turn is used to initialize the model texture. In the course of the procedure brackets will be placed onto individual teeth. As the difference between the iso-surface model and the reality increases, tracking will become less reliable. To account for these changes, we suggest updating and progressively refining a textured model of the jaw. This however is susceptible to the template update problem (i.e. drift) [10]. In combination with the edge-based approach we are able to exploit the accuracy of the edge-based registration with a quick but reliable pre-alignment using the textured model. The result is a robust alignment with a large capture range, despite challenging image data.

4   Experiments

For the evaluation of the proposed system, we created a prototype to study practical feasibility. In cooperation with orthodontic partners we acquired data for the real procedure of three teenage patients. The three videos each are about 20 minutes long and show the whole procedure from the perspective of a camera

mounted to the orthodontist's head. In each case DVT image data is available. In the following, we examine both the registration for single frames and the behavior for short video sequences.

4.1   Random Studies
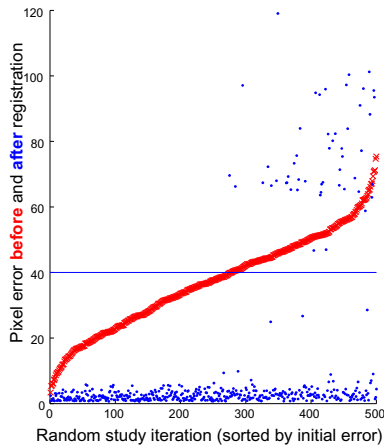
Fig. 4. Pixel error of a random study sorted by initial offsets

In a scenario where tracking is performed with the goal of a scene overlay, errors in image pixels are more relevant than in transformation parameter space. We therefore choose interest points on the surface of the teeth and compute the average projection errors. While 2D-3D point correspondences at significant edges and corners between the teeth were defined, they generally resulted in poor visual alignment due to the limited accuracy of placing those landmarks (i.e. target localization error TLE of over 5 pixels). Therefore, a quasi ground truth registration was defined based on the optimal visual alignment by the expert. In several random studies, we perturbed this ground truth alignment for all 6 pose parameters. The parameters were chosen, such that the x and y axis are parallel to the image plane and z is facing the camera with the origin at the center of the jaw. In Fig. 4 (left) we present a random study of 500 iterations as a typical representative. Translation was randomized in a range of ± 20 mm in x and y direction and ± 10 mm in z, while rotation in all three axes was randomized in a range of ± 10 degrees, enough to observe a failure of the algorithm in some instances. In the specific view of of the patient, similar to the one shown in Fig. 3 (left), this corresponds to an average pixel offset of 37.2 pixels on the 640×480 video image, which is well beyond expected inter-frame motion. After removal of 10% outliers, we were able to recover from an average error of 35.4 pixels to just 2.7 pixels, or an average of 2.1 mm and 4 degrees. In Fig. 4 (left) you can see the results of the random study including outliers, sorted by initial pixel error (red). Observe that the algorithm was successful in each case with offsets of less than 40 pixels, which is marked with a blue line in the plot, with the error after registration (blue) well below the red line. Even beyond that threshold, correct alignment is recovered in about 75% of the cases.

4.2   Image Sequences

We successfully tracked several sequences of all three patient videos. Visual alignment appeared accurate and reliable, especially around the incisors. Despite the dental prop, patients are moving their jaw during the procedure, which forced us to track upper and lower halves of the jaw independently. We believe that this
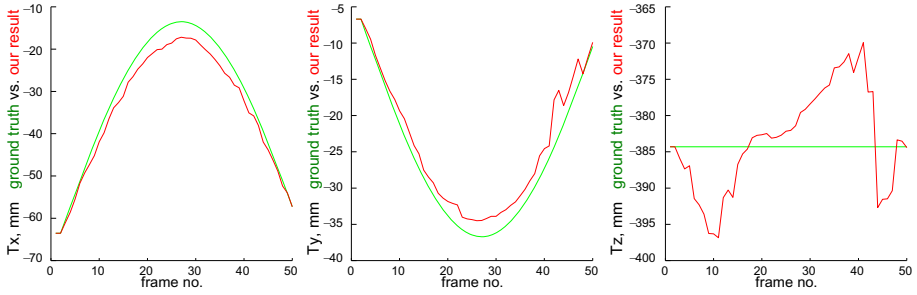
Fig. 5. Comparison of ground truth and our tracking results for a synthetic sequence

motion can be modeled with few parameters and included in one optimization, ultimately making alignment more stable, especially at the molars. Separation of mandible and maxilla can be performed automatically by fitting a plane.

In order to quantify behavior over a sequence of frames, we created a synthetic sequence using a textured model. While this experiment is simpler than real patient data, it allows us to compare the computed poses to ground truth. We chose a linear in-plane translation of x and y, as well as rotation about these axes and a motion returning to the starting position in 50 frames as a test case. See Fig. 5 (three plots on the left) for plots of the translation components of the resulting poses. Although the in-plane translation was recovered up to about 5 mm, there is an error in the z-translation by as much as 15 mm, which is expected for 2D-3D registration. As the z-axis is facing the viewer, translations in that direction have little effect on the image (and hence also on the final superimposition). The error in target points was 6.7 pixels average over the whole sequence.

5   Conclusion

We presented a tracking solution and novel guidance system for orthodontic correction. We focused on the feasibility of a tracking system based on a CT volume and the patient color video sequence. A multi-step algorithm was devised to use several aspects of the data. The proposed approach includes a dual iso-surface rendering method with distance based modulation to produce fast high-quality images of the gum line, paired with a textured model based pre-alignment and update step. In extensive random studies we could show correct registration of single images; more importantly, several sequences of real procedures were successfully tracked. In conclusion, we enabled a novel application of augmented reality in an orthodontics routine procedure. Future work could focus on a parameterization of jaw movement for concurrent tracking of both halves, as well as better handling of occlusion by the orthodontist's tool. While this work focused on recursive tracking for high accuracy, detection of the prop or the teeth could complement the current method (e.g. using an advanced approach such as

[4]). In the future our prototype has to be integrated with simulation and planning capabilities in order to create a fully practical solution, and a systematic quantitative evaluation of tracking accuracy performed.

References

1. Ammar, H.H., Ngan, P., Crout, R.J., Mucino, V.H., Mukdadi, O.M.: Three-dimensional modeling and finite element analysis in treatment planning for orthodontic tooth movement. American Journal of Orthodontics and Dentofacial Orthopedics 139, 59–71 (2011)
2. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. International Journal of Computer Vision 56(3), 221–255 (2004)
3. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 8, 679–698 (1986)
4. Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., Lepetit, V.: Gradient response maps for real-time detection of texture-less objects. IEEE Transactions on Pattern Analysis and Machine Intelligence (2011)
5. Klein, G., Murray, D.: Full-3D edge tracking with a particle filter. In: Proc. British Machine Vision Conference (BMVC 2006), vol. 3, pp. 1119–1128. BMVA, Edinburgh (September 2006)
6. Krüger, J., Schneider, J., Westermann, R.: ClearView: An interactive context preserving hotspot visualization technique. IEEE Transactions on Visualization and Computer Graphics 12(5) (September-October 2006)
7. Lepetit, V., Fua, P.: Monocular model-based 3D tracking of rigid objects: A survey. In: Foundations and Trends in Computer Graphics and Vision, pp. 1–89 (2005)
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)
9. Markelj, P., Tomazevic, D., Likar, B., Pernus, F.: A review of 3D/2D registration methods for image-guided interventions. Medical Image Analysis (2010)
10. Matthews, I., Ishikawa, T., Baker, S.: The template update problem. In: Proceedings of the British Machine Vision Conference (September 2003)
11. Mori, K., Deguchi, D., Sugiyama, J., Suenaga, Y., Toriwaki, J., Maurer Jr., C.R., Takabatake, H., Natori, H.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Medical Image Analysis 6(3), 321–336 (2002)
12. Wanschitz, F., Birkfellner, W., Figl, M., Patruta, S., Wagner, A., Watzinger, F., Yerit, K., Schicho, K., Hanel, R., Kainberger, F., Imhof, H., Bergmann, H., Ewers, R.: Computer-enhanced stereoscopic vision in a head-mounted display for oral implant surgery. Clinical Oral Implants Research 13, 610–616 (2002)

Intra-op Measurement of the Mechanical Axis Deviation: An Evaluation Study on 19 Human Cadaver Legs

Lejing Wang[1], Pascal Fallavollita[1], Alexander Brand[2], Okan Erat[1],
Simon Weidert[2], Peter-Helmut Thaller[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany

Abstract. The alignment of the lower limb in high tibial osteotomy (HTO) or total knee arthroplasty (TKA) must be determined intraoperatively. One way to do so is to deform the mechanical axis deviation (MAD), for which a tolerance measurement of 10mm is widely accepted. Many techniques are proposed in clinical practice such as visual inspection, cable method, grid with lead impregnated reference lines, or more recently, navigation systems. Each has their disadvantages including reliability of the MAD measurement, excess radiation, prolonged operation time, complicated setup and high cost. To alleviate such shortcomings, we propose a novel clinical protocol that allows quick and accurate intraoperative calculation of MAD. This is achieved by an X-ray stitching method requiring only three X-ray images placed into a panoramic image frame during the entire procedure. The method has been systematically analyzed in a simulation framework in order to investigate its accuracy and robustness. Furthermore, we validated our protocol via a preclinical study comprising 19 human cadaver legs. Four surgeons determined MAD measurements using our X-ray panorama and compared these values to a gold-standard CT-based technique. The maximum average MAD error was 3.5mm which shows great potential for the technique.

1 Introduction and Related Work

Patients with a condition known as knee osteoarthritis experience unfavorable wear and tear on the menisci and articular cartilage. These degenerative processes can cause the knee's protective tissues to wear on one side more than the other in a repetitive cycle of damage. A partial or total knee replacement can correct this condition when joint damage is beyond repair. A procedure known as a high tibial osteotomy (HTO), wedges open the upper tibia to reconfigure the knee joint. The weight-bearing part of the knee is shifted from degenerative or worn tissue onto healthier tissue. This procedure is typically reserved for younger patients with pain resulting from instability and malalignment. A high tibial osteotomy is generally considered to prolong the time before a total knee replacement, otherwise known as total knee arthroplasty (TKA), is required. During knee replacement surgery, the bone cuts and ligament balancing are

done in such a way as to produce a straightened, mechanically improved lower extremity. The challenge for surgeons in the operating room is to accurately and efficiently evaluate the mechanical axis of the lower extremity. The medial mechanical axis deviation (MAD), in which the mechanical axis of the lower extremity lies more than 10 mm medial to the knee joint center, is considered a failed procedure [1]. MAD is measured as the distance from the knee joint center to the line connecting the joint center of the hip and ankle (Figure 1(a)).

Recently, there has been increased interest in total knee arthroplasty with use of computer-assisted surgical (CAS) systems [2,3]. These systems are designed to increase the precision of implantation of the components. The studies have found a clear tendency toward improved alignment of the limb and the component position with use of computer-assisted surgical navigation. Another study indicated that there is no significant difference between TKA with use of computer-assisted surgical navigation and conventional TKA using X-ray [3]. Most of the complications are surgery related and not due to computer navigation except for the cases of system down or software failure. Nevertheless, limitations far exceeded the positive attributes, which include a learning curve for surgeons using the CAS systems, hefty cost, a cumbersome system setup including line of sight for tracking and on-site calibration, and difficult system registration [2,3]. Apart from using navigation systems that determine the axis, the most popular method in clinical practice is electrocautery cord and X-ray grid (Figure 1(b)), consisting of reference lines, which were proposed to verify the mechanical axis alignment [4,5]. With this process, a large number of X-ray images and consequently, exposure, are required to achieve recommended MAD measures. All necessary data for calculating the axis model is collected during the operation. Of critical importance, the joint centers must be located at the center of the X-ray image. This enables the surgeon to mentally assess and estimate lower limb axis alignment. The leg bones are repositioned and several other X-ray images are re-acquired to reassess axis alignment. This iterative process is repeated until there is surgeon satisfaction with respect to MAD tolerance. A postoperative radiograph is then acquired to verify a positive outcome of the surgery. In order to alleviate radiation exposure from these conventional methods, authors in [6] considered MRI, however they claim a significantly underestimated MAD measurement.

We propose a clinical protocol that has the potential to offer robust intraoperative assessment of MAD to the surgeon. Our protocol requires that the X-ray images cover the femoral head, the knee and the ankle for MAD measurements. These anatomical landmarks do not require central location in the image. The registration of the three X-ray images to the panorama is realized by an optical video camera attached to the C-arm fluoroscope gantry viewing a visual marker pattern underneath the operating table [7]. Thanks to a one-time calibration between the camera and the X-ray source [8], our system allows for real-time, radiation-free C-arm motion estimation. This low cost solution can be translated seamlessly in the current clinical setting and does not require additional hardware or calibration during surgery. The method has been systematically analyzed

(a)   (b)
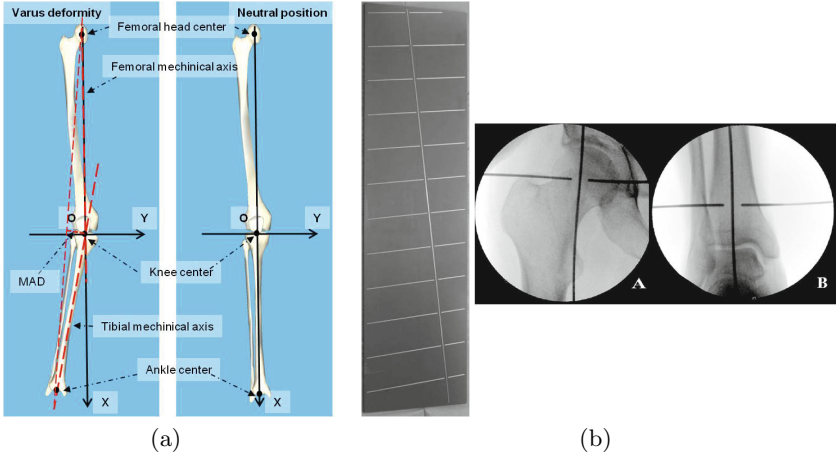
Fig. 1. (a) The mechanical axis deviation, MAD, is defined as the perpendicular distance between knee center and the line joining hip and ankle center. The left and right images show varus deformity and neutral position of the leg; (b) One standard in clinical practice is X-ray grid method for intraoperative axis alignment. The left image shows the X-ray grid board and the right image shows centres of the femoral head and of the ankle joint aligned with the reference lines of the board. Image courtesy of Liodakis et al. [9].

in a simulation framework in order to investigate its accuracy and robustness against potential error sources. we evaluate our new protocol by presenting results comparing the MAD measurements of 4 surgeons between ground-truth CT and X-ray panorama images obtained from 19 cadaver legs.

2   Methodology

2.1   X-ray Panorama Using Only 3 Images

The key step of image stitching is the estimation of the planar transformation for aligning images. Let a rotation $R_i \in \mathbb{R}^{3\times3}$ and a translation $t_i \in \mathbb{R}^3$ be defined from the coordinate system of the i-th camera view to the first camera view. In [10], the planar homography that aligns the i-th camera image to the first camera image is defined by

$$H_i = KR_iK^{-1} + \frac{1}{d_i}Kt_i n_i^T K^{-1} \tag{1}$$

where $K \in \mathbb{R}^{3\times3}$ is the intrinsic matrix of the camera. H_i is valid for all image points whose corresponding space points are on the same plane, i.e., stitching plane, defined by the normal vector $n_i \in \mathbb{R}^3$ and distance d_i to the origin in the coordinate system of the i-th camera view.

Metric measurements are possible for the 3D space plane whose plane parameters in the camera coordinate system are known [7]. The coordinate of a

3D space point $P \in \mathbb{R}^3$ in the first camera view can be derived from its image projection $m_i \in \mathbb{R}^3$ (expressed in homogeneous coordinates) of the i-th camera view by,

$$P = \frac{d_1 K^{-1}(KR_iK^{-1} + \frac{1}{t_i{}^T n_1 + d_1} Kt_i(R_i n_1)^T K^{-1})m_i}{n_1{}^T K^{-1}(KR_iK^{-1} + \frac{1}{t_i{}^T n_1 + d_1} Kt_i(R_i n_1)^T K^{-1})m_i} \tag{2}$$

With the estimated coordinates of the 3D points, the metric measurements between the points can be calculated.

2.2   The Limb Model Using Frontal Parallel Setup

We define the coordinate systems for the limb and C-arm, and then specify the frontal parallel setup for the limb and the C-arm. The limb coordinate system is first defined and its origin is at the knee center (see Figure 1(a)). Z-axis is the neutral AP vector for the front knee. The femoral mechanical axis is aligned with the X axis. Let the knee frontal plane (i.e. bone plane) be the XY plane. In the limb coordinate system, the knee center and femoral head center are fixed, while the ankle center locates differently on the XY plane depending on the hip-knee-ankle (HKA) angle. The coordinate system of the C-arm X-ray is expressed so that the origin is at the X-ray source center, the Z axis is pointing and perpendicular to the detector, and X and Y axes are along the image width and height (see Figure 2). Frontal parallel setup, i.e. the bone plane is parallel to the C-arm detector plane, defines the norm as $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ and distance as the translation along the Z axis of the C-arm. This setup is a commonly required clinical setup for enabling metric measurement in evaluating the frontal plane knee alignment [4,5] and X-ray image stitching [7,10]. It is usually verified by one control X-ray image showing the patella facing upwards [4,5].

2.3   Simulation Study

In the simulation study, let the C-arm coordinate system of acquiring the control X-ray image be the world coordinate system, whereby the control X-ray image defines the reference panorama frame. We set femur length (from the femoral head center to the knee center) and tibia length (from the knee center to the ankle center) to be 421mm and 355mm. We set HKA to be $5°$ and the true value of the corresponding MAD is accordingly 16.8 mm. The limb is positioned relative to the first C-arm view (the world coordinate system), such that it satisfies the frontal parallel setup. A frontal parallel setup constrains the norm of the frontal knee plane, i.e. no rotation around X and Y axes in the world coordinate system. In the frontal parallel setup, the distance of the bone plane to the origin is the translation along Z axis in the world coordinate system. The distance is unknown and can be arbitrary. General C-arm open space between the X-ray gantry and the detector is typically 1000 mm, and the patient is positioned close to the center nearer to the C-arm intensifier to achieve a larger field of view for
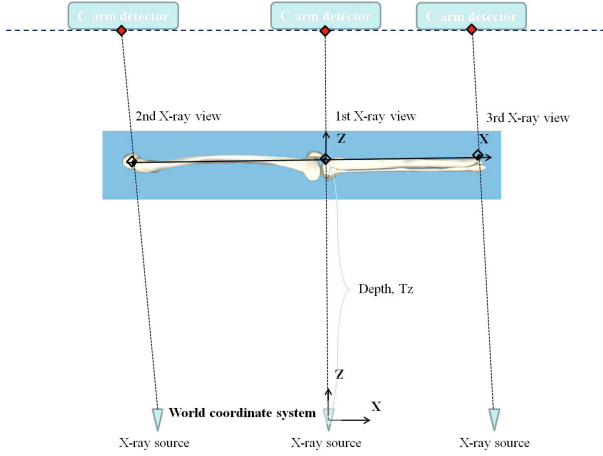
Fig. 2. Limb coordinate system definition. Its origin is at the knee center. Z axis is the neutral AP vector for the front knee. The femoral mechanical axis is aligned with the X axis. The knee frontal plane (i.e. bone plane) is defined be the XY plane.

the surgeon. For this reason, we choose 700mm for the distance. We initially set rotations and translation about XYZ to be zero. The above parameter settings fix the transformation from the limb to the world coordinate system, which defines the frontal parallel setup used in our simulation. Measurement errors are always a possibility during a typical surgery primarily due to the distance of the bone plane not being accurately calculated or not being in the frontal parallel setup. To correct for these [7] require a manual estimation of the distance between the bone plane and marker pattern plane. Alternatively, [10] proposed to align the contours of the bone plane in two X-ray images for recovering the measurement error caused by the incorrect distance. We choose the former option as our solution does not require the overlapping area between the X-ray images in order to further reduce radiation exposure.

For clarity, we use the term estimated setup to represent the frontal parallel setup, since it is not the true actual setup. We make estimated setup constant in our simulation study, since most clinical methods for panorama generation and measurements are based on the assumption of frontal parallel setup. The actual pose of the limb is obtained by adding deviations to the estimated setup. To simulate the image projections of the three anatomical points in the individual images and the C-arm motions between image acquisitions, the actual pose of the limb is applied. To stitch the images (i.e. computing the coordinates of the imaged points in the panorama frame) and perform the measurements, we use the estimated pose of the limb. For each of following simulation experiments, we perform 1000 trails to compute the difference between the estimated and true MAD values as the final MAD errors.

2.4   Preclinical Study

19 human cadaver legs were used and manually bent to introduce a random valgus or varus misalignment while keeping the knee fully extended. Each individual leg was positioned in a wooden box and placed on a carbon table with a visual marker pattern of 2378 x 1682 mm rigidly attached underneath it. The leg positioning was such that the patella faced upwards (i.e. confirmed by an X-ray image). The Siemens Powermobile C-arm (Erlangen, Germany) was translated along the carbon table in a PA gantry position for acquiring the three X-ray images. Three X-ray images of femoral head, the knee and the ankle were acquired and subsequently stitched together. A CT scan of the leg was conducted immediately after (i) to ensure that we have the same leg position between X-ray stitching and CT, and (ii) as ground truth reference comparison. Four clinicians with different experience levels (i.e. 2 expert surgeons, 1 resident surgeon, and 1 last year medical student) were involved in our study. The anatomical landmarks defining the mechanical axis and the knee center were manually determined by each observer individually for all of the CT images and panoramic X-ray images in order to allow MAD calculation. For the CT measurement we used the K-Pacs Software v.1.6.0.

3   Results and Discussion

A 10 mm MAD value is a typical range of the healthy patients [1]. In clinical practice, a successful operation of lower limb correction is confirmed by a 0mm of MAD after operation. Therefore, we use 10mm as the MAD error tolerance to ensure the actual value of MAD is less than 10mm when its measured value is 0mm.

Simulation Study Results: We simulate the inaccurate frontal parallel setup by adding random errors within $\pm 5°$ to the rotations around XYZ axes of the limb pose. The mean MAD error calculated is: 0.52 ± 0.51 mm. The image coordinates of the anatomical points are manually selected by surgeons in clinical practice. The incorrect determination of the image coordinates, depending on experience of the surgeon, is simulated by adding random errors within ± 10 pixels to the true image coordinates of hip-knee-ankle. The mean MAD error calculated is: 1.56 ± 1.07 mm. To assess the influence of the inaccurate distance of the bone plane on the MAD measurements, we add distance errors to ruler readings between -100 mm to 100 mm with the increments of 20 mm to the true distance. For each distance error, we compute MAD errors using 1000 trails with and without applying random errors of the rotations and image coordinates. Figure 3 shows the results and indicates that it is important to accurately estimate distance of bone to source.

Preclinical Results: CT-based measurement is regarded as gold-standard. The change in measured MAD values between CT and the proposed method is the MAD error. First a t-test was performed; using CT data MAD values of 14.75 ± 9.16 mm were achieved compared to 13.25 ± 7.66 mm when using the C-arm
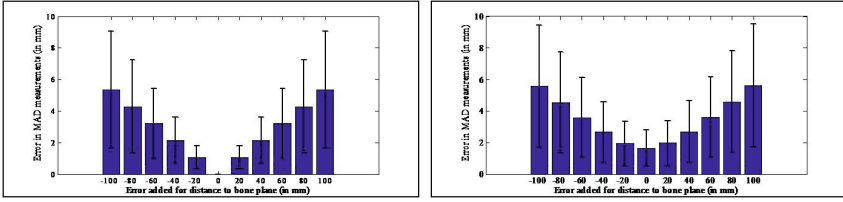
Fig. 3. Average and standard deviation MAD errors, in millimeters, for (Left), distance errors on ruler reading alone, and (Right) both distance errors on rulers coupled with pixel errors in the X-ray images

stitching data. From this, there is no significant difference in the measured data ($p > 0.1$). Secondly, we performed a Pearson Correlation that displayed a highly significant correlation between CT and our method ($r = 0.96$). We performed an error analysis of 100 trials for each of the 19 cadaver legs by calculating the mean and standard deviations of mechanical axis deviation when adding random errors described in Figure 4. Results for each participant compared to ground-truth CT show that the MAD error is well below the recommendation of 10mm [1] for HTO and TKA procedures guaranteeing axis alignment. The MAD errors between all the examiners are below 1mm. The student is slightly better. This is another proof that the proposed method is reliable against inter-observers. Our maximum 3.5mm MAD error shows our method robustness. Note that Leg alignment involves multiple MAD measurements. For the entire procedure, the effect of our proposed MAD measurement method will be multiplied. As agreed upon by our three expert surgeons, an average of 8-12 X-ray images are typically acquired for one MAD measurement compared to our 3.
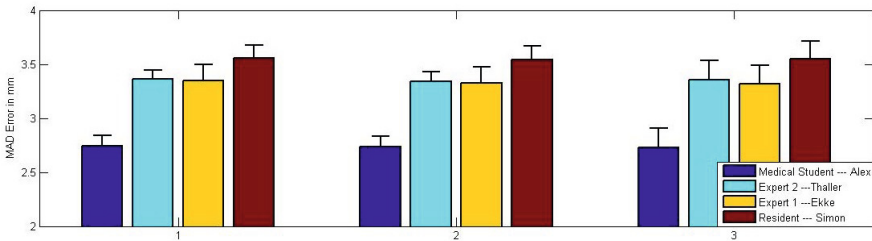
Fig. 4. An average MAD error of at most 3.5 mm for the mechanical axis deviation compared to ground-truth CT shows the robustness of our protocol. 1- errors only added to hip-ankle-knee pixel coordinates in X-ray; 2- errors added to ruler height; 3- the combination of both errors in X-ray and ruler.

4   Conclusion

Existing methods and solutions for MAD intraoperative calculation demand a high level of surgeon experience since both hip and ankle centers have to be

brought on a common grid or reference line and be located in the center of the X-ray [4,5]. This results in a complex procedure and many X-ray shots. Our proposed technique requires only three X-ray images to be acquired showing the hip center, the knee center and the ankle center, thereby reducing procedural time. Via a one-time ruler measurement of the distance between marker pattern and knee, we always achieve accurate intraoperative MAD measurements enabling our protocol to become a potential alternative to standard clinical practice. Furthermore, our protocol can be generalized to any standard C-arm worldwide as long as the C-arm pose is available. We await ethics board approval for a patient study evaluation.

References

1. Paley, D.: Principles of Deformity Correction, 2003 edn. Springer (April 2003)
2. Kim, Y., Kim, J., Yoon, S.: Alignment and orientation of the components in total knee replacement with and without navigation support: a prospective, randomised study. The Journal of Bone and Joint Surgery 89(4), 471–476 (2007)
3. Kim, Y., Kim, J., Choi, Y., Kwon, O.: Computer-assisted surgical navigation does not improve the alignment and orientation of the components in total knee arthroplasty. The Journal of Bone and Joint Surgery 91(1), 14–19 (2009)
4. Saleh, M., Harriman, P., Edwards, D.J.: A radiological method for producing precise limb alignment. The Journal of Bone and Joint Surgery 73(3), 515–516 (1991)
5. Krettek, C., Miclau, T., Grn, O., Schandelmaier, P., Tscherne, H.: Intraoperative control of axes, rotation and length in femoral and tibial fractures. Technical Note 29(suppl.3), 29–39 (1998)
6. Liodakis, E., Kenawey, M., Doxastaki, I., Krettek, C., Haasper, C., Hankemeier, S.: Upright MRI measurement of mechanical axis and frontal plane alignment as a new technique: a comparative study with weight bearing full length radiographs. Skeletal Radiology 40(7), 885–889 (2011)
7. Wang, L., Traub, J., Weidert, S., Heining, S.M., Euler, E., Navab, N.: Parallax-free intra-operative x-ray image stitching. Med. Imag. Anal. 14(5), 674–686 (2010)
8. Navab, N., Heining, S.M., Traub, J.: Camera augmented mobile c-arm (camc): Calibration, accuracy study and clinical applications. IEEE Trans. Med. Imag. 29, 1412–1423 (2010)
9. Liodakis, E., Kenawey, M., Liodaki, E., Mommsen, P., Krettek, C., Hankemeier, S.: The axis-board: An alternative to the cable technique for intraoperative assessment of lower limb alignment. Technology and Health Care 18(3), 165–171 (2010)
10. Yaniv, Z., Joskowicz, L.: Long bone panoramas from fluoroscopic x-ray images. IEEE Transactions on Medical Imaging 23(1), 26–35 (2004)

Real-Time Quantitative Elasticity Imaging of Deep Tissue Using Free-Hand Conventional Ultrasound

Ali Baghani[1], Hani Eskandari[1], Weiqi Wang[1],
Daniel Da Costa[1], Mohamed Nabil Lathiff[1], Ramin Sahebjavaher[1],
Septimiu Salcudean[1], and Robert Rohling[1,2]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
baghani@ece.ubc.ca
[2] Co-appointed with the Department of Mechanical Engineering, University of British Columbia, Vancouver, BC, Canada

Abstract. In this article an ultrasound elastography technology is reported which provides quantitative images of tissue elasticity from deep soft tissue. The technique is analogous to Magnetic Resonance Elastography in the use of external mechanical vibrations which can penetrate deep tissue. Multifrequency steady-state mechanical vibrations are applied to the tissue at the skin and tissue displacements are measured by a conventional ultrasound system. Absolute values of tissue elasticity are computed in real-time for each frequency and displayed to the physician. The quantitative elasticity images produced by the technology are validated with magnetic resonance elastography images as the gold standard on standard elasticity phantoms. Preliminary *in-vivo* data from healthy volunteers are presented which show the potential of the technology for clinical use. The system is currently being used in different clinical studies to image kidney fibrosis, liver fibrosis, and prostate cancer.

1 Introduction

Elastography has emerged as an imaging modality providing new information to the clinician about the mechanical properties of tissue [1]. It has found a place in the imaging of breast lesions [2], liver fibrosis [3], and is being investigated in many other clinical areas such as targeting of prostate cancer for biopsy and focal therapy [4].

The first generation of elastography technology was developed on ultrasound machines [5] and created images of relative elasticity of tissue. The clinician applied a manual compression with the probe to deform the tissue, and the elastography system measured the tissue strain. Under certain assumptions, the tissue strain is inversely proportional to the tissue stiffness, and therefore strain images can show the contrast in tissue stiffness. The first generation elastography

is now available on many medium to high-end ultrasound systems under different brand names such as elastography, strain imaging, real-time elastography, etc.

The general trend in radiology towards quantitative imaging, together with the difficulties in applying the manual compression, called for more innovation and through the efforts of different groups, second generation elastography technologies were born [6,7]. A second generation elastography technique creates quantitative elasticity images where the contrast in the image is the absolute elasticity of the tissue.

Elastography has been developed primarily with ultrasound or magnetic resonance imaging [8] as the underlying imaging device to track tissue displacements. To cause tissue displacements, different methods have been devised with mechanical transient excitation [2], mechanical steady-state excitation [8,9], and acoustic radiation force (shear wave) transient excitation [6,7] to name a few.

The majority of the magnetic resonance imaging techniques use a steady-state mechanical excitation to image the tissue elasticity [8, 10, 11]. The idea is to measure the wavelength of the steady-state wave patterns in the tissue from which the wave speed can be estimated. The wave speed depends on the mechanical properties of tissue, and is generally higher in stiffer tissue compared to softer tissue.

The reported technology uses analogous techniques as used in magnetic resonance elastography for ultrasound elastography. A steady-state mechanical vibration is applied to the tissue while the tissue is imaged by the ultrasound. From the sequence of ultrasound images, the tissue displacements and wave patterns are computed. The local wavelength of the wave pattern is then estimated to create a map of the tissue stiffness which is displayed in real-time.

This article gives an overview of the technology and reports the most recent advances. In particular we report the first direct comparison of the technique with MRE on a standard quality assurance phantom. Steady-state excitation was first used in sonoelasticity to image tissue stiffness [12]. It has also been used before to produce MRE-type elasticity images [13]. However this is the first report of an implementation for "real-time" operation with "free-hand" conventional ultrasound. Two novel ideas which have enabled these advances are the use of a "thin-slice" consisting of a few planes for 3D data acquisition, and the fast implementation of all the image processing pipeline on a graphics processing unit (GPU). These advances are reported for the first time in this article. Based on these qualities, the technology holds promise for ultrasound guided procedures, such as biopsies, by providing additional quantitative information to the clinician.

2   Methods

The system has been implemented on a SonixTouch platform (Ultrasonix Medical Corp., Richmond, BC, Canada) (Fig. 1 (a)). Two prototype systems have been developed based on the Texo and Ulterius software development kits. A mechanical vibration source (LDS V203, LDS/B&K, Norcross, GA) has been

mounted on the flexible arm of the SonixTouch. The hardware of the system has been upgraded with a CUDA supporting graphics card (GTX 580 GPU, NVIDIA, Santa Clara, CA). The SonixTouch collects the RF-data which are then transferred to the memory of the graphics card. All the computational algorithms for absolute elastography have been implemented in CUDA and run on the GPU in real-time. The absolute elasticity images are transferred back to the CPU memory and displayed.

(a) (b)
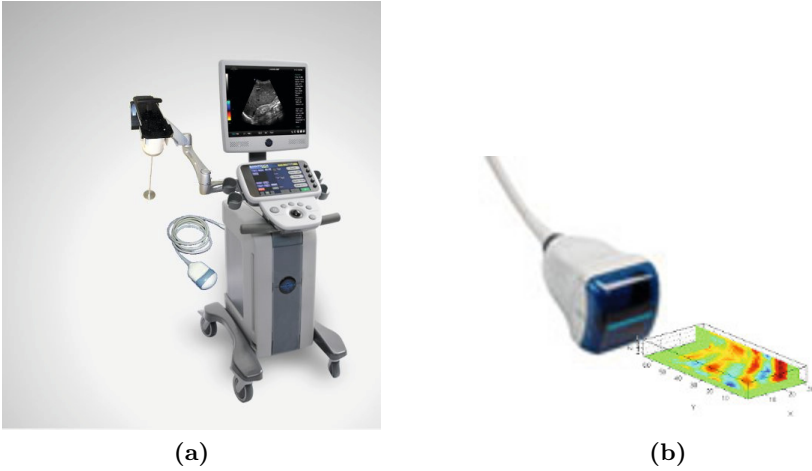
Fig. 1. (a) A photo of the actual elastography system with the vibration source mounted on the SonixTouch arm (b)The 4DL14-5 3D probe and the thin volume of displacement data acquired from 7 planes

The image processing pipeline of the system is shown in Fig. 2. The RF-data are processed by the speckle tracking block to find the time-domain displacements, which are then converted to frequency-domain displacements or phasors by a Fourier transform for each frequency in the excitation. The phasors are then passed through the inversion algorithm. An algorithm used for the speckle tracking is described in [14] and different algorithms for the inversion are compared in [15]. Here we used the Local Frequency Estimation method [8]. Average computational times are compared in table 1 for CPU vs. GPU on test data sets.

The system can be used with 2D and 3D probes to perform absolute elastography. With 2D probes, the variations of the waves in the elevational direction cannot be observed and measured. As a result, the elasticity images obtained with 2D probes suffer some over-estimation in the elasticity values (10 to 60
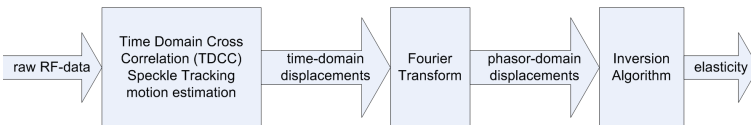
Fig. 2. Image processing pipe-line

Table 1. Comparison of computational time between CPU and GPU implementations of the elastography algorithms

	CPU time (ms) Core i5 660@3.33GHZ	GPU time (ms) GTX 480
Speckle Tracking	1600	50
Fourier Transform	110	9
Inversion	1800	12

percent is typical). The error has some dependency on the placement of the vibration source. The major benefit of 2D probes, however, is the faster acquisition of US data which results in a higher frame rate.

The elastograms obtained with 3D probes are more accurate and robust, but they suffer from a lower frame rate. A thin volume (typically consisting of 5 to 7 planes) is acquired to capture the variation of the wave in the elevational direction (see Fig. 1 (b)). Currently the system generates elasticities at a rate of 1 image every 2 seconds with a 4DL14-5 3D probe and 80mm depth of imaging.

Fig. 3 shows variation of the wave pattern (real part of the displacement phasor) in the elevational direction as measured by a 3D probe in two different configurations. By comparing the two cases, it can be observed that by changing the placement of the excitation source relative to the probe, the variation of the displacement in the elevational direction can be minimized. In this case a relatively accurate estimate of the elasticity can be obtained by using a 2D probe.

(a)   **(b)**
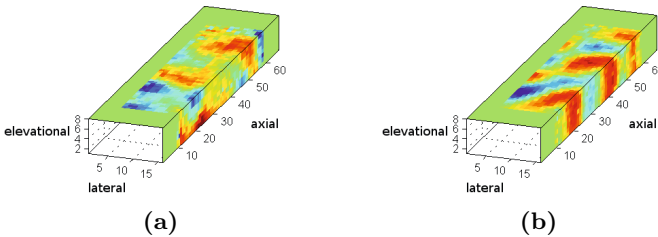
Fig. 3. The variation of displacement in the elevational direction depends on the placement of the excitation source (a) more variation (b) less variation in the elevational direction

3   Validation with MRE

The mechanism of creating waves in the tissue by steady-state mechanical vibration is analogous in the reported system and MRE. The methods which are used to invert the wave patterns and compute the elasticities are also analogous. Therefore, one would expect that an absolute elastogram produced by the reported system should have the same appearance and values as an MRE absolute elastogram, as a gold standard.

To prove this hypothesis, experiments were performed with a standard elasticity phantom; The elasticity quality assurance phantom model 049 (CIRS Inc, Norfolk, VA, USA) was imaged using a 3T Achieva MRI system (Philips Inc, Netherlands). This phantom includes different spherical inclusions of varying stiffness and sizes. The frequency of excitation was 200Hz. The elastogram obtained for a horizontal plane passing through the larger inclusions is shown in Fig 4(a).

The same phantom was imaged using a 4DL14-5 linear 3D probe (Ultrasonix Medical Corp., Richmond, BC, Canada) with the system. The probe was mounted on a linear stage and moved along the long axis of the phantom to image a larger area covering multiple inclusions. The vibration was applied with the arm mounted voice coil of the system at 200Hz. At each location of the linear stage, a slim volume of ultrasound data was acquired and used to reconstruct a single plane of elasticity (mid-plane of the slim volume). By moving the 3D probe to a new position, another elasticity plane was produced and so forth. By putting these single plane elasticity images together a whole volume elasticity map was produced. The result is shown in Fig. 4(b). A comparison of Figs 4(a) and 4(b) provides some evidence for the validity of hypothesis. The elasticity values obtained from the two methods are reported in table 2 as well as the manufacturer values.
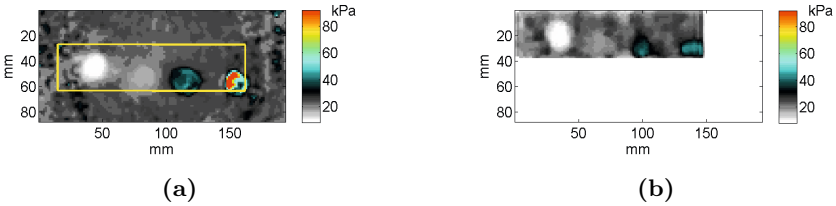
Fig. 4. (a) Magnetic Resonance Elastography and (b) Ultrasound elastography images acquired by the system

Table 2. Comparison of average elasticity values between MRE, ultrasound elastography, and manufacturer values

	MRE (kPa)	US/E (kPa)	Man. (kPa)
Softest Inclusion	8	9.5	6
Soft Inclusion	17	17	17
Hard Inclusion	42	40	54
Hardest Inclusion	54	45	62
Background	24	24	29

4 *In Vivo* Results

The system has been used to image different tissue *in-vivo* and *ex-vivo* with different probes. In this section, a sample of the data collected so far is reported.

The 4DL14-5 linear 3D transducer was used to image the liver of consenting healthy volunteers *in-vivo* to a depth of 80mm. The transducer was placed on the side of the patient and images were obtained intercostally. A multi-frequency excitation, consisting of 50, 80, and 100 Hz components was applied via the external vibrator. As with MRE studies, we observe that the elasticity increases as a function of frequency [16]. The trend can be seen in the elastograms of a volunteer, at the three frequencies, shown in Fig. 5. In all the images, a stiffer layer corresponding to intercostal muscle can be observed near the skin, while the liver has a fairly homogeneous appearance in the elastogram.
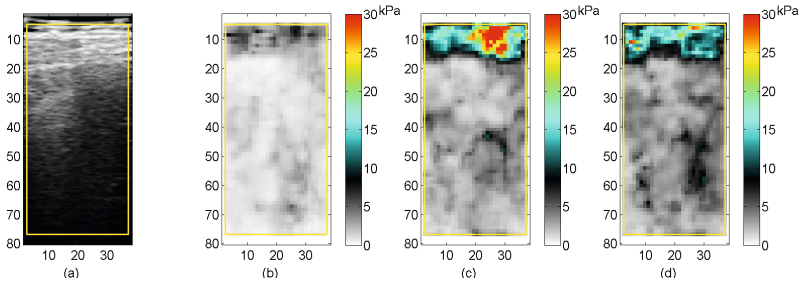
Fig. 5. The frequency dependence of elasticity from a healthy volunteer liver (a) B-mode image (b) elastogram at 50 Hz, (c) elastogram at 80 Hz, (d) elastogram at 100 Hz
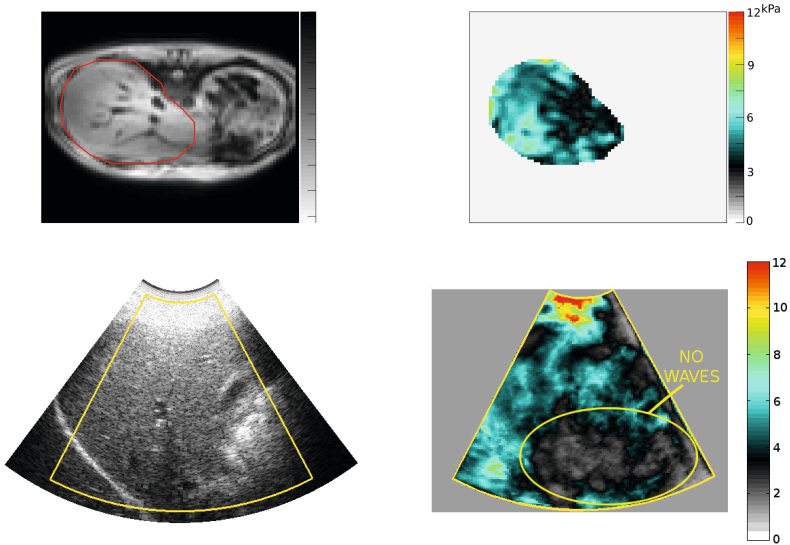
Fig. 6. Top: MRI image and MRE image at 56 Hz, Bottom: B-Mode and elastogram from a healthy volunteer liver averaged between 55 and 60 Hz, obtained with a low frequency curved 3D transducer

In another experiment, a comparison was made between MRE and the proposed system. The 4DC7-3 curved transducer was used to image the liver of a consenting healthy volunteer at 55 Hz and 60 Hz down to a depth of 150mm. The transducer was placed on the side of the volunteer and images were obtained intercostally. The B-mode and average elasticity images are shown in Fig. 6 as well as the MRE data of the same volunteer collected at 56 Hz. The range of values observed are similar in both elastograms. However, a more carefully designed experiment is needed with registration of the ultrasound and MRI images before a conclusion can be made.

5   Conclusion

In this article an ultrasound elastography technology is reported which is analogous to magnetic resonance elastography in the type of excitation and inversion algorithms used. The absolute elasticity values obtained by the system are validated against MRE values as a gold standard on a quality assurance elasticity phantom. At low frequencies (30-100 Hz) the mechanical vibrations can penetrate deep tissue, and provide the necessary displacements for measurement of tissue elasticity at depths of 150mm. Typical *in vivo* elastograms obtained by imaging healthy volunteer livers with different transducers are reported. The values reported are similar to the values reported in the literature from MRE studies.

References

1. Lerner, R., Parker, K., Holen, J., Gramiak, R., Waag, R.: Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets. Acoustical Imaging 16, 317–327 (1988)
2. Bercoff, J., Chaffai, S., Tanter, M., Sandrin, L., Catheline, S., Fink, M., Gennisson, J.L., Meunier, M.: In vivo breast tumor detection using transient elastography. Ultrasound Med. Biol. 29(10), 1387–1396 (2003)
3. Huwart, L., Peeters, F., Sinkus, R., Annet, L., Salameh, N., ter Beek, L.C., Horsmans, Y., Beers, B.E.V.: Liver fibrosis: Non-invasive assessment with MR elastography. NMR in Biomedicine 19, 173–179 (2006)
4. Lorenz, A., Sommerfeld, H., Garcia-Schurmann, M., Philippou, S., Senge, T., Ermert, H.: A new system for the acquisition of ultrasonic multicompression strain images of the human prostate in vivo. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 46(5), 1147–1154 (1999)
5. Ophir, J., Cespedes, I., Ponnekanti, H., Yazdi, Y., Li, X.: Elastography: A quantitative method for imaging of elasticity of biological tissues. Ultrasonic Imaging 13, 111–134 (1991)
6. Bercoff, J., Tanter, M., Fink, M.: Supersonic shear imaging: A new technique for soft tissue elasticity mapping. IEEE Trans. Ultrason. Ferroelect. Freq. Contr. 51(4), 396–409 (2004)
7. Nightingale, K., Palmeri, M., Nightingale, R., Trahey, G.: On the feasibility of remote palpation using acoustic radiation force. J. Acoust. Soc. Am. 110(1), 625–634 (2001)

8. Muthupillai, R., Lomas, D., Rossman, P., Greenleaf, J., Manduca, A., Ehman, R.: Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. Science 269(5232), 1854–1857 (1995)
9. Baghani, A., Brant, A., Salcudean, T., Rohling, R.: A high frame rate ultrasound system for the study of tissue motions. IEEE Trans. Ultrason. Ferroelect. Freq. Contr. 57(7), 1535–1547 (2010)
10. Sinkus, R., Tanter, M., Xydeas, T., Catheline, S., Bercoff, J., Fink, M.: Viscoelastic shear properties of in vivo breast lesions measured by MR elastography. Magnetic Resonance Imaging 23, 159–165 (2005)
11. Asbach, P., Klatt, D., Hamhaber, U., Braun, J., Somasundaram, R., Hamm, B., Sack, I.: Assessment of liver viscoelasticity using multifrequency MR elastography. Magnetic Resonance in Medicine 60(2), 373–379 (2008)
12. Lerner, R., Huang, S., Parker, K.: Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues. Ultrasound in Medicine and Biology 16(3), 231–239 (1990)
13. Muller, M., Gennisson, J., Deffieux, T., Sinkus, R., Annic, P., Montaldo, G., Tanter, M., Fink, M.: Full 3D inversion of the viscoelasticity wave propagation problem for 3D ultrasound elastography in breast cancer diagnosis. In: IEEE Ultrasonics Symposium, pp. 672–675. IEEE (2007)
14. Zahiri-Azar, R., Salcudean, S.: Motion estimation in ultrasound images using time domain cross correlation with prior estimates. IEEE Transactions on Biomedical Engineering 53(10), 1990–2000 (2006)
15. Baghani, A., Salcudean, S., Honarvar, M., Sahebjavaher, R., Rohling, R., Sinkus, R.: Travelling wave expansion: A model fitting approach to the inverse problem of elasticity reconstruction. IEEE Transactions on Medical Imaging 30(8), 1555–1565 (2011)
16. Asbach, P., Klatt, D., Schlosser, B., Biermer, M., Muche, M., Rieger, A., Loddenkemper, C., Somasundaram, R., Berg, T., Hamm, B., et al.: Viscoelasticity-based staging of hepatic fibrosis with multifrequency MR elastography. Radiology 257(1), 80–86 (2010)

A Comparative Study of Correspondence-Search Algorithms in MIS Images

Gustavo A. Puerto and Gian-Luca Mariottini

Department of Computer Science and Engineering,
University of Texas at Arlington, 416 Yates Street, 76019 Texas

Abstract. The ability to find image similarities (*feature matching*) between laparoscopic views is essential in many robotic-assisted Minimally-Invasive Surgery (MIS) applications. Differently from feature tracking methods, feature matching does not make any restrictive assumption about the sequential nature of the two images or about the organ motion, and could then be used, e.g., to recover tracked features that were lost due to a prolonged occlusion, a sudden endoscopic-camera retraction, or a strong illumination change. This paper provides researchers in the medical-imaging computing community with an extensive comparison of the most up-to-date feature-matching algorithms over a large (and annotated) data set of 100 MIS-image pairs obtained from real interventions. The *accuracy* of these methods, as well as their ability to *consistently retrieve* as many good matches as possible, are evaluated for popular feature detectors. In addition, the dataset and the software implementations of these methods are made freely available on the Internet.

1 Introduction

In robotic-assisted Minimally-Invasive Surgery (MIS), the ability to find image similarities (*feature matching*) between laparoscopic views of the same scene is an essential need in many applications, such as shape recovery [1,2], calibration [3], structure and camera-motion estimation [4,5], augmented reality [6,7], etc.

Thus far, some algorithms have addressed the similarity-search problem in MIS video sequences only in the case that the two views are consecutive frames of a video[1], or when the organ motion is periodic [8,9].

On the contrary, *feature-matching algorithms* [10–13] make very limited (or no) assumptions about the observed scene, or about the two given images. These algorithms have thus the potential to become a core-component in each of the aforementioned MIS applications, as they could be used to *recover tracked features* lost due to a prolonged occlusion, a sudden endoscopic-camera retraction, or a strong illumination change.

It is then of primary importance to provide researchers in the medical-imaging computing community with a comparative study of the *accuracy* of each of the aforementioned algorithms, and to evaluate their ability to *consistently retrieve*

---

[1] This problem is usually known as *feature tracking*.

as many good matches as possible. Existing works in the computer-vision field have thus far examined only the performance of *feature detection* [14–16], while other works have compared matching strategies in MIS [17, 18] based only on *local appearance* of descriptors, which are not robust to ambiguities.

In this work we present an extensive comparison of the aforementioned state-of-the-art feature-matching techniques [10–13] over a large (and annotated) data set of 100 MIS-image pairs obtained from real interventions. In particular, we measure both the ability of the methods to establish correct correspondences, as well as their accuracy in a tracking-recovery scenario. To the best of our knowledge, this is the first work that thoroughly compares and analyzes feature-matching algorithms; in addition, we make available on the Internet the dataset and the software implementations of these methods.[2]

The paper is organized as follows: Sect. 2 provides a brief overview and discussion of the above feature-matching algorithms. Sect. 3 illustrates the results of our extensive experimental evaluation and discusses the algorithms' performance. Finally, Sect. 4 draws conclusions and illustrates future extensions.

2   Overview of Feature-Matching Algorithms

We consider a pair of images, $\mathcal{I}^t$ (*training*, e.g., before occlusion) and $\mathcal{I}^q$ (*query*, e.g., after occlusion), and two corresponding sets of *image features* (e.g., SIFT [10], SURF [14], ASIFT [19], etc.) $\mathcal{F}^t$ and $\mathcal{F}^q$, extracted from $\mathcal{I}^t$ and $\mathcal{I}^q$, respectively. Each image feature consists of a *keypoint* [10] (with pixel position, $\mathbf{x} \triangleq [x, y]^T$, scale σ, and orientation θ), and of a *descriptor vector*, which captures the local appearance around the keypoint position.

The main goal of a *feature-matching algorithm* is to retrieve pairs of similar features (*correspondences*) among two images. Existing strategies, first find a set of *initial* (or candidate) *matches*, $\mathcal{M}$, by using appearance-based criteria that relies on the similarity between descriptors [15, 17]. Among the most common criteria: simple thresholding, Nearest Neighbor (NN) distance ratio, etc.

Existing feature-matching algorithms mostly differ for their geometric-constraint phase, which is adopted to reduce the number of ambiguous matches originated by image regions with similar appearance.

Lowe's [10]:   This method estimates a *single* Affine Transformation (AT) [20], $\mathbf{A}$, which maps keypoints, $\mathbf{x}^t$ and $\mathbf{x}^q$, from $\mathcal{I}^t$ to $\mathcal{I}^q$.

Among the initial matches, $\mathcal{M}$, only those that obey to such a geometric AT will be considered as *inliers*, $\mathcal{M}_{in}$. In particular, these *inliers* consist of those matches with a low symmetric keypoint reprojection error, e_s. For a pair of generic points, $\mathbf{u}^t \in \mathcal{I}^t$ and $\mathbf{u}^q \in \mathcal{I}^q$, e_s is defined as follows[3],

$$e_s(\mathbf{u}^t, \mathbf{u}^q, \mathbf{A}) = 0.5\|\tilde{\mathbf{u}}^q - \mathbf{A}\tilde{\mathbf{u}}^t\|_2 + 0.5\|\mathbf{A}^{-1}\tilde{\mathbf{u}}^q - \tilde{\mathbf{u}}^t\|_2. \tag{1}$$

---

[2] Web: http://ranger.uta.edu/%7egianluca/feature%5fmatching.

[3] $\tilde{\mathbf{u}}$ represents the extension to homogeneous coordinates of $\mathbf{u}$.

Discussion: The approach in [10] works well for matching features on *planar objects*. However, since the majority of features in a MIS scenario lie on non-planar (or deformable) surfaces, this method can only recover few matches. For example, from our experimental results, we observed that the refined matches (inliers) that better support **A** will only be those at the intersection of the object surface and of its (single) tangent plane. The other matches, distributed over the rest of the surface, will be considered as outliers.

Adaptive Multi-Affine (AMA) [11]: AMA relaxes the assumption of [10] by estimating a set of *multiple* ATs. Each AT is associated to a cluster of matches; as a result, the inliers (matches) detected by AMA will be distributed along the entire organ's surface. In AMA a set of clusters are first estimated as in [10], which are then sent to a cascade of RANSAC-based affine estimators. For each cluster, this estimation phase computes those features supporting a single AT $\mathbf{A}_i$. These clusters are then adaptively quantized by *k-means*, and each transformation is estimated, together with their inliers $\mathcal{M}_i$.

Discussion: AMA extracts more inliers than Lowe's approach. The final inliers in the image are spatially distributed over the entire organ's surface.

Agglomerative Correspondence Clustering (ACC) [12]: ACC determines the set of refined matches by employing a hierarchical clustering algorithm based on an agglomerative (bottom-up) strategy. This strategy iteratively merges pairs of matches (or clusters of matches) into a single cluster based on a dissimilarity measure between matches (or clusters). This dissimilarity measure consists of *both* geometric and appearance constraints. Finally, the algorithm iteratively merges clusters according to both their dissimilarity measure and a linkage criteria, to generate the final clusters.

Discussion: ACC automatically detects similar portions in the two images without requiring previous knowledge of the scene. However, ACC requires the user to specify a larger number of (non-intuitive) parameters than [10,11].

Del Bimbo's Local Homography Mapping (DBA) [13]: DBA generates a set of *n synthetic images*, $\mathcal{I}_i^t, i = 1, \ldots, n$, obtained by applying to $\mathcal{I}^t$ a set of known random homographies $\mathbf{H}_i^t$ [20]. From the set, $\mathcal{F}'$, of all the extracted features, a geometric constraint is used to remove repeated ones. The homography $\mathbf{H}_i$ and the associated inliers are then estimated by weighting each synthetic homography according to the remaining features. A final appearance-based phase (e.g., NCC) is used to removes the remaining outliers.

Discussion: DBA requires a large number of random (thus uncontrolled) synthetic images. As a result, DBA can only detect a low number of matches.

3 Experimental Results and Discussion

We compared the performances of the above algorithms (Sect. 2), in two scenarios: *i)* a highly controlled in-lab test with a non-planar object, and *ii)* a large

MIS image dataset (100 image pairs) acquired from six surgical videos (with many cases of occlusions, large camera motion, specular reflections, etc.)

In each scenario, the performances of each algorithm have been assessed by measuring both the efficiency in detecting correct matches (*matching performance*), and the accuracy in mapping corresponding points between image pairs (*mapping performance*). For each image pair, a set of manually-labelled matches has been used to evaluate the *matching performance* by comparing the *Receiver Operating Characteristic* (ROC) curves [21], which depict the relative trade-off between the *sensitivity* (algorithm's capability to successfully retrieve correct matches, i.e., benefits), and the *1-specificity* (i.e., algorithm's proficiency to erroneously retrieve wrong matches, i.e, costs). The *mapping performance* is measured by using a set of manually-labelled (*ground-truth*) corresponding points, and by measuring statistics on the symmetric reprojection error [20].

Similarly to [12,15,17,18], our database consists of a set of manually-labelled (ground-truth) SIFT matches captured by an expert user. Each match was manually labeled as 'correct' or 'wrong' by carefully observing their position on the two images (note that only the most certain matches were labelled as 'correct'). In addition, our database contains a set of manually-selected correspondences between each image pair. Our database also includes sets of initial SURF and ASIFT matches, however they are not labeled due to the large number of their initial matches and the associated complex labelling process[4]. Note that we have chosen to avoid to limit our datasets only to the strongest SURF and ASIFT features in order not to alter the performance (e.g., the percentage of correct matches) of these feature detectors.

Note that, for a fair comparison, all the aforementioned algorithms use the same sets of initial matches obtained by using NN distance ratio with threshold values[5] in the range of 0.8 and 0.9. The parameters and thresholds of each algorithm were chosen to maximize efficiency. In particular, we used a threshold for the affine model of 5 pixels for both AMA and Lowe; a cutoff value of 5 was chosen for the ACC's linkage function, and a normalized cross correlation coefficient of 0.7 for DBA.

3.1   In-Lab Experiments

Two views of a non-planar object were used as a highly-controlled in-lab experiment (resolution 640×480). The initial 354 SIFT matches (220 were labelled as correct), while the SURF, and the ASIFT initial matches are 330, and 4903, respectively. In addition, 41 corresponding corners were manually selected.

Fig. 1 shows the ROC curves for each of the four algorithms parametrized by their score values[6]. In Lowe and AMA the scores represent the negative

---

[4] For example, ASIFT can produce more than 4000 initial matches per image, thus rendering the ground-truth (manual) labelling almost unfeasible.

[5] These thresholds were chosen to maximize the number of initial matches. This choice does not affect our analysis since the initial matches are common to all the methods.

[6] Algorithm's quantitative measure for the classification task.

symmetric-reprojection errors; for ACC the scores are the dissimilarity measure of the matches, which is also based on the symmetric reprojection error. For DBA, the scores represent the absolute value of the normalized cross-correlation coefficients. The square markers in Fig. 1(a), represent different score values, $\{5, 20, 50, 70\}$ for Lowe, AMA, ACC, and $\{0.9, 0.8, 0.7, 0.5\}$ for DBA. Fig. 1(b) depicts a more focused comparison of the sensitivity values for Lowe, AMA and ACC[7]. The marked points have each an associated score value.
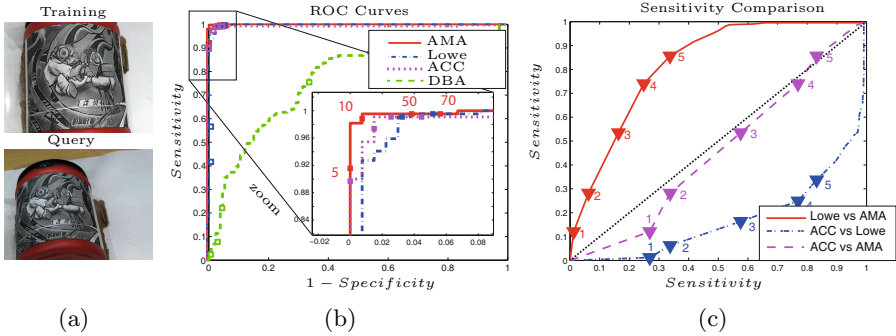
Fig. 1. *In-lab experiment - Matching performance*: (a) Query and Training images; (b) ROC curves of the four algorithms. AMA, ACC and Lowe are very competitive, being AMA slightly superior (c.f., inset); (b) Sensitivity comparison.

Table 1. *In-lab experiment - Mapping performance*

Alg.	Avg.Err. SIFT (pix)	Time SIFT	Sen/1-Sp SIFT	Avg.Err. SURF (pix)	Time SURF	%Inl. SURF	Avg.Err. ASIFT (pix)	Time ASIFT	%Inl. ASIFT
Lowe	22.1±23.7	0.22	0.28/0	20.8±22.5	0.16	17.3	19.3±21.0	5.1	15.3
DBA	32.9±64.7	120	0.22/0.045	17.4±27.2	661	21.5	20.2±50.4	721	25.9
AMA	5.06±6.09	1.26	0.98/0.001	5.35±7.73	1.48	60.0	2.62±2.81	36.1	65.3
ACC	6.33±8.9	0.5	0.96/0.015	7.46±8.13	0.52	33.9	4.58±4.96	433	42.3

Table 1 provides statistical results for the *mapping performance*. The even columns show the mean and standard deviation of the symmetric reprojection error. The odd columns show the (SIFT) Sensitivity/1-specificity of Lowe, AMA, ACC (less than 5 pixels) and DBA (correlation coefficient greater than 0.7). For SURF and ASIFT, these columns contain the percentage of inliers with respect to the total number of matches, which indicate the detection power of each algorithm.

Discussion: The ROC curves in Fig. 1(a) show that Lowe, AMA and ACC, differently from DBA, have high *matching performance*. However, for the same threshold of 5 pixels, AMA has better performance than ACC and Lowe (c.f., Fig. 1(b)). An increase in the threshold value will provide an improved sensitivity for Lowe, AMA, and ACC, without compromising the 1-specificity.

---

[7] We did not include a comparison over 1-specificity because of space limits.

The results in Table 1 show that AMA has superior *mapping performances* than the other methods, when using either SIFT, SURF, and ASIFT. This indicates that AMA (multiple ATs) can consistently detect as inliers many features lying on non-planar organs' surface. ACC achieves comparable results than AMA. However, AMA is computationally more expensive, than Lowe, and ACC (except when ASIFT features are used). Lowe (single AT) is the fastest method, but can only capture few matches. ACC exhibits a higher computational complexity when a large number of features is used (e.g., ASIFT), and DBA shows poor mapping performances caused by the random (thus uncontrolled generation of the synthetic images. We also observed that the mapping accuracy cannot be improved by simply increasing the AT-fitting thresholds for Lowe and AMA, since this could make the algorithms more sensitive to outliers.

3.2   Surgical-Images Dataset

Our MIS dataset consists of 100 images with a resolution of 704×480 pixels. The set of ground-truth correspondences has an average of 20 points. For SIFT features, the corresponding *average* ROC curves for each algorithm are illustrated in Fig. 2(a). The vertical lines indicate the confidence intervals of the mean with respect to different score values[8], with a significance level of 95%. Fig. 2(b) shows the comparison of the sensitivity among pairs of (the most competitive) methods.
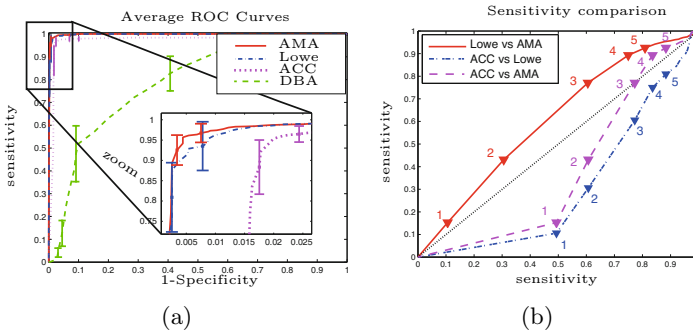
Fig. 2. *MIS dataset - matching performance*: (a) Average ROC curves. The vertical lines show the 95% confidence intervals. (b) Comparison of the algorithms' sensitivity.

Table 2 summarizes the average results for the algorithms' mapping performance for fixed threshold and for different types of features.

Discussion: AMA shows superiority in both metrics, followed by Lowe and ACC, with similar performances (see inset), while DBA is again far behind due to its incapability to filter out image similarities present in the MIS images. Observe that Lowe is significantly the fastest algorithm. Note that ACC's is faster than AMA and DBA for SIFT and SURF, while it becomes dramatically slower than

---

[8] Same scores values are used: $\{5, 20, 50, 70\}$, and $\{0.9, 0.8, 0.7, 0.5\}$.

AMA with ASIFT. Differently from Lowe and ACC, AMA achieves good performances within the given threshold (5 pixels). AMA's best-mapping performance is attained when ASIFT features are used. We also observed that AMA cannot reliably detect isolated matches in the case of textureless regions.

ACC has similar sensitivity than AMA, but at a slightly larger 1-specificity. We observed that ACC dissimilarity function, which is not fully geometrical, is more flexible than Lowe's or AMA constraints. This flexibility allows ACC to capture matches even if they do not satisfy the affine map, but have similar appearance. However, this can produce an increased number of false positives and, thus, a larger mapping error. In addition, ACC has a large number of (non-intuitive) parameters to tune.

Table 2. *MIS dataset: mapping performance*

Alg.	Avg.Err. SIFT	Time SIFT	Sen/1-Sp. SIFT	Avg.Err. SURF	Time SURF	%Inl. SURF	Avg.Err. ASIFT	Time ASIFT	%Inl. ASIFT
Lowe	3.8±3.6	0.3±0.1	0.72/.001	3.6±2	0.2±0.2	39.6	3.6±3.4	2.5±1.2	42.3
DBA	21±37	122±14	0.52/.098	23.3±30.3	406±62	25.1	21.6±30	519±23	38.3
AMA	3.1±3.1	1.0±0.6	0.9/.002	3±3.1	1.2±1.1	59.4	3.3±3.1	13.9±13.5	54.5
ACC	5.3±4.7	0.3±0.1	0.61/.015	7.1±8	0.3±0.3	59.0	10.4±13	140±228	53.5

Fig. 3 shows an example of the matching performance of the four algorithms for given thresholds. Note that DBA has problems to filter out the ambiguities; Lowe and ACC obtain similar results, being AMA capable of retrieving a larger number of matches (covering most of the organ's surface).
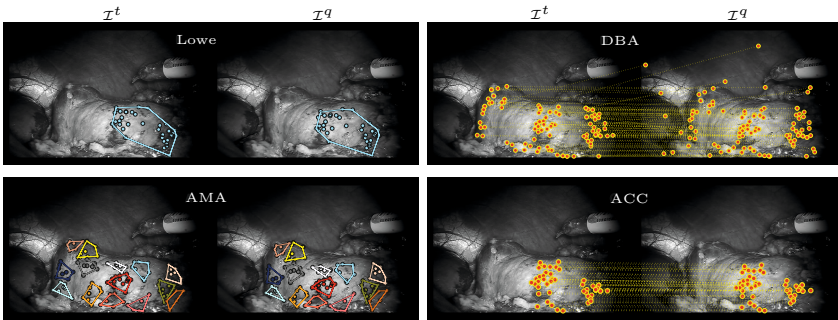
Fig. 3. Matching performances (better seen in color). Lowe only detects few matches limited to a portion of the organ. AMA, ACC and DBA retrieve matches distributed over all the organ. ACC and DBA are more sensitive to ambiguities.

4   Conclusions

This work presented for the first time a comparison of four state-of-the-art feature-matching algorithms. Differently from existing works, the four methods were compared among several types of popular feature descriptors, by using

a large and manually-annotated MIS data set. We compared their performance based on two metrics to assess each algorithm's proficiency to detect correct correspondences, as well as the accuracy in mapping corners between images. Lowe, AMA, and Cho's exhibit competitive results, while AMA has superior performance at for the same threshold levels, because of its better adaptation to non-planar surfaces and higher robustness to outliers. The software implementations of these algorithms and our dataset are made freely available on the Internet. As future work we plan to add more types of features, and to increase the MIS dataset.

References

1. Visentini-Scarzanella, M., Mylonas, G.P., Stoyanov, D., Yang, G.-Z.: *i*-BRUSH: A Gaze-Contingent Virtual Paintbrush for Dense 3D Reconstruction in Robotic Assisted Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 353–360. Springer, Heidelberg (2009)
2. Totz, J., Mountney, P., Stoyanov, D., Yang, G.-Z.: Dense Surface Reconstruction for Enhanced Navigation in MIS. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 89–96. Springer, Heidelberg (2011)
3. Stoyanov, D., Darzi, A., Yang, G.Z.: Laparoscope Self-calibration for Robotic Assisted Minimally Invasive Surgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 114–121. Springer, Heidelberg (2005)
4. Hu, M., Penney, G.P., Rueckert, D., Edwards, P.J., Bello, F., Casula, R., Figl, M., Hawkes, D.J.: Non-rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 34–42. Springer, Heidelberg (2009)
5. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking. IEEE Sig. Proc. Mag. 27, 14–24 (2010)
6. Su, L.M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D.: Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3d-ct to stereoscopic video registration. Urology 73(4), 896–900 (2009)
7. Ukimura, O., Gill, I.S.: Image-fusion, augmented reality, and predictive surgical navigation. Urol. Clin. of North America 36(2), 115–123 (2009)
8. Stoyanov, D., Mylonas, G., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
9. Mountney, P., Yang, G.Z.: Motion Compensated SLAM for Image Guided Surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010)
10. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comp. Vis. 60(2), 91–110 (2004)
11. Puerto, G., Adibi, M., Cadeddu, J., Mariottini, G.: Adaptive multi-affine (AMA) feature-matching algorithm and its application to minimally-invasive surgery images. In: Proc. IEEE/RSJ Int. Conf. Intel. Rob. Syst., pp. 2371–2376 (2011)

12. Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: Proc. 9th Int. Conf. Comp. Vis., pp. 1280–1287 (September 2009)
13. Del Bimbo, A., Franco, F., Pernici, F.: Local shape estimation from a single keypoint. In: Proc. Comp. Vis. Patt. Rec. Workshops, pp. 23–28 (June 2010)
14. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Proc. European Comp Vis. Conf., pp. 404–417 (2006)
15. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. 60, 1615–1630 (2005)
16. Gauglitz, S., Höllerer, T., Turk, M.: Evaluation of interest point detectors and feature descriptors for visual tracking. Int. J. Comp. Vis. 94(3), 335–360 (2011)
17. Mountney, P., Lo, B., Thiemjarus, S., Stoyanov, D., Zhong-Yang, G.: A Probabilistic Framework for Tracking Deformable Soft Tissue in Minimally Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 34–41. Springer, Heidelberg (2007)
18. Giannarou, S., Visentini-Scarzanella, M., Yang, G.Z.: Affine-invariant anisotropic detector for soft tissue tracking in minimally invasive surgery. In: Proc. 6th IEEE Int. Conf. Symp. Biom. Imag., pp. 1059–1062 (July 2009)
19. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. J. on Imag. Sciences 2(2), 438–469 (2009)
20. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge Univ. Press (2000)
21. Fawcett, T.: An introduction to ROC analysis. Patt. Rec. Letters 27(8), 861–874 (2006)

3D Reconstruction in Laparoscopy with Close-Range Photometric Stereo

Toby Collins and Adrien Bartoli

ALCoV-ISIT
CNRS and Université d'Auvergne, Clermont-Ferrand, France
{Toby.Collins,Adrien.Bartoli}@gmail.com

Abstract. In this paper we present the first solution to 3D reconstruction in monocular laparoscopy using methods based on Photometric Stereo (PS). Our main contributions are to provide the new theory and practical solutions to successfully apply PS in close-range imaging conditions. We are specifically motivated by a solution with minimal hardware modification to existing laparoscopes. In fact the only physical modification we make is to adjust the colour of the laparoscope's illumination via three colour filters placed at its tip. Once calibrated, our approach can compute 3D from a single image, does not require correspondence estimation, and computes absolute depth densely. We demonstrate the potential of our approach with ground truth ex-vivo and in-vivo experimentation.

1 Introduction

An important computer vision task in Minimally Invasive Surgery (MIS) is to recover the 3D structure of organs and tissues viewed in endoscopic images and videos. A general solution to this has many important applications, including enhanced intra-operative surgical guidance, depth perception, 3D motion estimation and compensation, novel-view synthesis and improving pre-operative/intra-operative data registration. In the literature, the main practical monocular reconstruction approaches so far are based on Structure-from-Motion (SfM). However, since this is correspondence based, it is error prone and at textureless regions 3D cannot be recovered. SfM also requires very strong assumptions on surface motion (e.g. rigid or periodic motion), and requires sufficient motion baseline. By contrast, PS offers a very different approach for 3D which is based on photometric constraints using three or more light sources [2,16,9]. PS is attractive since it provides dense 3D estimates, does not require correspondence estimation, and can compute 3D from just a single colour image. However, to date PS has not been applied to 3D reconstruction in MIS. Our main contributions are to provide the theory and practical solutions to successfully apply PS to the very close-range imaging conditions of MIS. In this paper we focus on laparoscopy. On the hardware side, our solution takes a standard monocular laparoscope, modified only with three colour filters (red, green and blue) placed at its tip. This corresponds to a practical and very inexpensive modification. The physical dimensions remain unchanged and it does not require any strobing or synchronised triggering between the camera and light source.
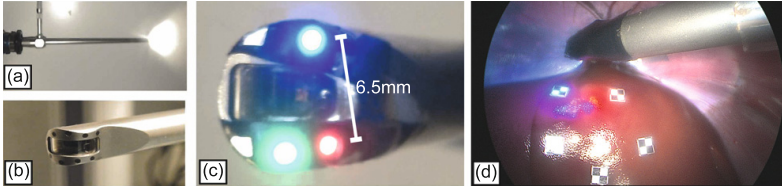
Fig. 1. Modification made to a standard laparoscope (a,b) to facilitate practical in-vivo photometric stereo by fitting three colour filters at its tip (c). Photograph of in-vivo tests (d).

3D Reconstruction in MIS. Several different research directions for 3D reconstruction in MIS have emerged over the recent years. These differ in the sensing hardware required to compute 3D. At one end of the spectrum are dedicated 3D sensing devices. These have included Structured Light (SL) setups [1] and Time-of-Flight (ToF) cameras [12]. SL requires additional instruments, which may clutter the scene and to date, neither SL nor ToF sensors have proved sufficiently reliable in practice. Stereo endoscopes have also been tried for 3D reconstruction [14,4,8]. While promising, these are limited by fixed camera convergence angles and stereo baseline, and perform poorly at textureless regions. At the other end of the spectrum are passive monocular methods. These require no additional instruments and compute 3D using the raw video feed. This however is an extremely challenging computer vision problem. Some progress has been made using SfM and Simultaneous Localisation And Mapping (SLAM) [3,6,11]. These have been tried in several domains including reconstructing the abdominal cavity [5] and heart [7]. However standard SfM and SLAM assume the 3D scene is rigid, which is unrealistic during intervention. Nonrigid-SLAM extensions have been proposed, yet these require strong motion models, such as cyclic deformation [11], learned low-rank shape bases [7] or conformal surface extension [10]. Shape-from-Shading (SfS) is another passive method tried [13,17] that exploits the relationship between geometry, pixel intensity and scene illumination. In contrast to SfM it can return dense 3D, but it currently has major limitations. These include the inability to handle surface discontinuities, and inherent unreliability due to SfS being a very weakly constrained problem.

3D Reconstruction with Distant Light Photometric Stereo. PS can be considered the generalisation of SfS to multiple light sources. In prior work, the *distant light source model* is nearly always used [2,16,9]. This serves as a basis for us, but is unsuitable at close-range where illumination attenuation is significant. A given point $\mathbf{q} = (u, v)$ in an image projects out into 3D according to an (unknown) depth function $z(u, v) : \mathbb{R}^2 \to \mathbb{R}^+$. Its 3D position is given by $\mathbf{p} = \mathbf{K}^{-1} (u, v, 1)^\top z(u, v)$, where $\mathbf{K}$ denotes the camera's perspective intrinsics (which implies image distortion effects have been undone.) It is assumed that $\mathbf{p}$ is lit by $K \geq 3$ lights whose directions are given by the vectors $\mathbf{l}_k$. For an RGB camera, we have effectively three light sensors, with each channel sensitive to different parts of the light spectrum. Denote $c_i(u, v)$ to be the radiometrically

corrected image intensity of the pixel for the ith colour channel. In standard distant-light PS, 3D shape at $\mathbf{p}$ is constrained by lambertian reflectance according to: $c_i(u,v) = \sum_k \mathbf{l}_k \cdot n(u,v) \mathbf{A}_{ik}$. Here, $n(u,v)$ denotes the surface normal. $\mathbf{A}$ is a $3 \times k$ matrix where $\mathbf{A}_{ik} \geq 0$ holds the illumination response of the i^{rh} channel as a function of surface albedo and the kth light's power spectrum [2]. Distant Light PS involves using these constraints to solve for $n(u,v)$. This is a small, quadratically-constrained Linear Least Squares (LLS) problem. Once estimated, dense 3D shape is recovered by integrating $n(u,v)$ in a second optimisation phase. Note however that absolute depth is not recoverable, and shape is given up to an unknown scale factor.

2   Close-Range Photometric Stereo

In this section we generalise the PS problem to handle close range light conditions. We present a new low-parameter illumination model which models very well a laparoscope's light source and give a method for quick and practical light calibration. We retain the lambertian model in this work, and handle specularities via saturation detection. This simplified model allows for tractable dense 3D reconstruction. We further advocate lambertian constraints in another important respect. By placing polarizing filters over the light and camera, specular reflection can be hugely reduced, leaving mostly only the lambertian term. Thus, with filters, the lambertian model is arguably a good one to use (for reconstruction purposes) even if the viewed surface comprises specular reflections. We start by extending the PS constraints to the following general form:

$$c_i(u,v) = l_k(p(u,v)) \cdot n(u,v) w_k(p(u,v))\mathbf{A}_{ik}, \quad p(u,v) = \mathbf{K}^{-1}(u,v,1)^\top z(u,v) \quad (1)$$

Here $l_k(\cdot) : \mathbb{R}^3 \to \mathbb{R}^3$ is now not a constant, but a *spatially varying* light vector function. $w_k(\cdot) \in [0:1]$ is also a spatially varying function that gives the amount of light attenuation from the kth light source to a point in 3D space. We say that the light model is *calibrated* if $l_k(\cdot)$ and $w_k(\cdot)$ are known. Close-range PS then involves solving the following variational least squares system:

$$\underset{z(u,v)}{\arg\min} \int_{(u,v)\in\Omega} \sum_{k=1}^{K} \sum_{i=1}^{3} \left(l_k(p(u,v)) \cdot n(u,v)w_k(p(u,v))\mathbf{A}_{ik} - c_i(u,v)\right)^2 + \\ \lambda \int_{(u,v)\in\Omega} \|\nabla z(u,v)\|_2^2 \quad (2)$$

Here the domain Ω denotes an image region bounded by the surface. The first line enforces the PS constraints, and the second enforces surface smoothness weighted by λ. Let us now step back and compare close-range PS to the distant-light case. Firstly (2) cannot be broken down into two convex problems (normal estimation, followed by depth estimation). This is because the PS constraints depend on both depths and normals. As such, it is a harder optimisation problem. However, it is the fact that (2) depends on depths that allows us to recover *absolute* distances to the camera (in mm), unlike distant-light PS.

2.1   Illumination Modelling

We now turn to modelling and calibrating the light source functions $l_k(\cdot)$ and $w_k(\cdot)$. Our goal is to have accurate models, yet which can be calibrated easily and optimally. We propose using an attenuating point light source, with a bivariate polynomial which can model light fall-off caused by both distance and angular attenuation. This is a flexible model and a generalisation of the inverse-squared fall-off model [17], which we have found to be rather poor. The model's parameters comprise firstly a light source centre: $\mathbf{u}_k \in \mathbb{R}^3$. The illumination vector at any 3D point $\mathbf{p}$ is given by the unit direction $l_k(\mathbf{p}) = (\mathbf{p} - \mathbf{u}_k)/\|\mathbf{p} - \mathbf{u}_k\|_2$. The attenuation function is a joint function of the distance from $\mathbf{p}$ to $\mathbf{u}_k$: $d(\mathbf{p}, \mathbf{u}_k) = \|\mathbf{p} - \mathbf{u}_k\|_2$, and the angular attenuation w.r.t. the light's *principal direction* $\mathbf{v}_k$: $\psi(\mathbf{p}) = \angle(l_k(\mathbf{p}), \mathbf{v}_k)$. This angular attenuation is important to model the spotlight characteristics of the light source. Here $\angle(\cdot \cdot)$ denotes taking the angle between two vectors. The attenuation function then writes as:

$$w_k^{-1}(\mathbf{p}) = \sum_{s=1}^{S} \sum_{t=1}^{T} \mathbf{W}_{st}^k d(\mathbf{p}, \mathbf{u}_k)^s \psi(\mathbf{p}, \mathbf{v}_k)^t \qquad (3)$$

Here $\mathbf{W}^k$ holds the kth light's polynomial coefficients up to order (S, T).

2.2   Light Calibration

Light calibration involves finding, for each coloured light source, the values $\{\mathbf{u}_k, \mathbf{v}_k, \mathbf{W}_k\}$. Typically for endoscopes the light sources remain rigidly fixed in the camera's coordinate frame, which means that calibration can be done in a one-time offline process. We divide the calibration problem into first determining the light centres $\{\mathbf{u}_k\}$, and then using these to determine the attenuation terms $\{\mathbf{v}_k, \mathbf{W}_k\}$. This 2-stage approach gives a convex solution to light calibration, and so global optimality is guaranteed. The light centres $\mathbf{u}_k$ can be found easily by detecting and triangulating their positions on a reflective calibration target [15] (Fig. 2(a)). To calibrate $\{\mathbf{v}_k, \mathbf{W}_k\}$ we use Eq. (1) to optimise these terms using ground-truth training samples. The data is acquired using images of a diffuse planar calibration target with a checker pattern printed on one side (Fig. 2(b)). The pattern gives us the plane's pose in each image. For each colour filter, we gather a large set of training samples $\{(c_r, c_g, c_b, \mathbf{n}, \mathbf{p})\}$. Now, for a given value of $\mathbf{v}_k$, $\mathbf{W}_k$ can be optimised via LLS. We thus calibrate by densely sampling $\mathbf{v}_k$ over the unit sphere, solving for $\mathbf{W}_k$, and retaining the solution with minimal least-squares error w.r.t. Eq. (1). We can select the best (S, T) by minimising the fitting error on a separate validation set. In practice it is usually unnecessary to go beyond 4th order.

2.3   Reflectance Model Learning

For any tissue we wish to recover we also need an estimate of $\mathbf{A}$. For $K = 3$ lights this is only a small 3×3 matrix and can be determined with training data. There are two main approaches one could take for this. The first is to learn $\mathbf{A}$ prior to intervention for a range of tissue classes. The second is to assume the training
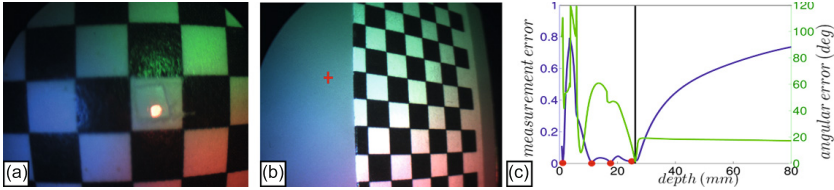
Fig. 2. Calibration of light source centres using specular reflection (a) and light source attenuation using views of a lambertian planar target (b). For the pixel marked at the cross, there may exist multiple depths which locally satisfy the close-range PS constraints.

data can be acquired during intervention by some other means. Currently we adopt this second approach, and place a small marker on the tissue (e.g. Fig. 4 (a-e)). By computing the marker's pose, and sampling tissue intenities close to the marker, we have the necessary data to compute **A** from Eq. (1).

3 3D Reconstruction

The 3D reconstruction problem involves solving (2) which is a challenging non-convex problem. Here we present an effective 2-stage approach to find a good minimum. In stage 1 we first solve for depth at each pixel *locally* using only that pixel's colour information. When computed in isolation from other pixels these solutions are however usually non-unique. In stage 2 these local solutions are then resolved by solving depth *globally* across the image. At any pixel (u, v), the intensities $c_r(u, v), c_g(u, v), c_b(u, v)$ provide us with 3 constraints on shape according to Eq. (1). Locally, we have three unknowns, one for $z(u, v)$ and two for $n(u, v)$. This is a polynomial optimisation problem whose number of solutions depends foremost on the order of $w_k(\cdot)$. We propose a fast method to find these solutions as follows. We regularly sample depth in the range $\tilde{z}(u, v) \in [0 : z_{max}]$, where z_{max} denotes the maximum working distance of the laparoscope (typically 150mm). Using Eq. (1), each sample is used to solve for a putative surface normal $\tilde{n}(u, v)$, which is a small convex sub-problem. We can then evaluate the solution pair $(\tilde{n}(u, v), \tilde{z}(u, v))$ against the measured intensities predicted by Eq. (1), and retain the solutions which are optimal. We illustrate this approach in Fig. 2(b-c). In Fig. 2(b) a planar surface is 34.5mm from the camera's optical centre. For the pixel marked at the cross, we show in 2(c) on the x-axis the depth along the pixel's viewing ray. In green we show the angular error of the surface normal estimated at that depth using Eq. (1). In blue we plot the prediction error of the pixel's intensity. Clearly, there is a 4-fold solution ambiguity, marked by red dots with zero prediction error. The rightmost solution is closest to the true depth, marked by a black line.

The sets of solutions computed at each pixel can be resolved in a second process by enforcing surface continuity between pixels. This can be achieved by constructing a Markov Random Field (MRF), whose nodes correspond to pixels and edges connect neighbouring pixels. These edges enforce consistency

between pixels' normal and depth estimates, and form a graph with sub-modular pairwise interaction terms. We have found the MRF's energy can be minimised well with belief propagation and the solution gives us a reasonable initial solution to the depth map. This is then refined by optimising the original problem (2). In practice we discretised Ω using the pixel grid and iteratively refine $z(u, v)$ with nonlinear Gauss-Newton iterations.

4   Experimental Results

Ex-Vivo Experimentation. We have tested our approach ex-vivo using two organs; a section of pig liver and a pig kidney. We have performed ground truth evaluation by first scanning these surfaces with a Structured Light Scanner (SLS). In Fig. 3(a,g) we show the kidney and liver ground truth surfaces. To learn the organs' reflectance models, we attached a small planar checker marker to the organ to give us depth and normal information (Fig. 3(b,h)), and used the non-specular tissue colour around the marker as the intensity training data. We handled external laparoscope tracking using a mounted calibration target, giving us the coordinate transform from the laparoscope's view to the 3D SLS surface. We then imaged the organs with the laparoscope at varying positions (Fig. 3(c,d,i,j)). For each image, we manually segmented the organ from the background to obtain Ω. In these experiments we did not use polarizing filters and specularities were handled with simple methods by detection based on colour saturation. For any specular region, its pixel data does not contribute to the first term in Eq. (2). 3D reconstruction was then performed. With our current Matlab implementation this takes approximately 30s to process an image. However, much can be parallelised so a GPU implementation would be significantly faster. We used the same value of λ for all images (which was set by hand) and measured the absolute error in depth against ground truth. We show the results for the four images below in Fig. 3(e,f,k,l). In general the surfaces are reconstructed quite faithfully. Greater errors occur towards some boundaries of the surface, which is due to surface inter-reflection from the background and slight mis-alignments of the ground truth scan. Note in Fig. 3(l) the larger error occurs at a region where the red channel becomes saturated, which corresponds to losing a PS constraint.

In-Vivo Experimentation. We have also obtained some preliminary in-vivo experimental results by testing reconstruction on the liver of a live pig under anesthetic. To acquire ground truth data our surgeon placed 5 4.5mm wide checker-markers on the liver using surgical graspers[1]. This gives a sparse set of ground truth depths and surface normals. Fig. 4(a-e) show a selection of images of the markers taken by the laparoscope. We used one of these markers to learn in-vivo the liver's surface reflectance model by sampling pixel intensities at tissue locally surrounding the marker. To perform reconstruction, we took the image domains Ω to be the entire image, but excluding the marker locations.

---

[1] We thank Dr. Revaz Botchorishvili for his kind help in acquiring the in-vivo data.
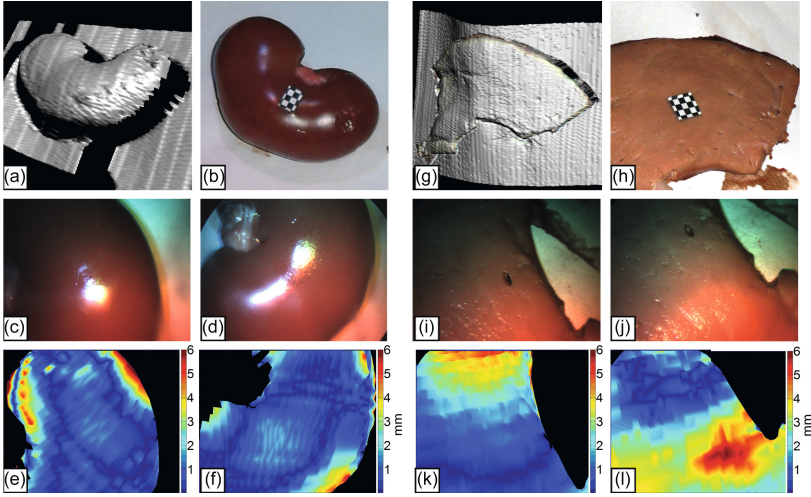
Fig. 3. Ex-vivo experimental validation of close-range PS
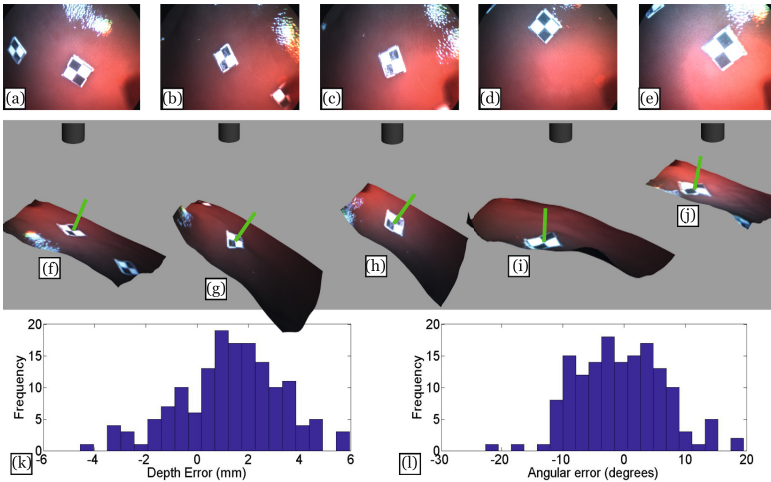
Fig. 4. In-vivo experimental validation of close-range PS

The 3D reconstructions are shown in Fig. 4(f-j), each corresponding to the input image shown above it. We also render the laparoscope's tip, indicating its absolute distance to the surface. In green we mark the predicted surface normal at each marker, computed from the gradient of the reconstructed surface. In total, we have performed reconstruction for 30 images. Quantitative performance has been studied by measuring the error in the predicted depths and normals of the markers. In Fig. 4(k-l) we show the error distribution in both depth (in mm) and surface orientation (in degrees).

5   Conclusion and Future Work

In this paper we have aimed to answer an important open question: can PS be used to successfully reconstruct surfaces in the close-range conditions of MIS? Our preliminary results suggest yes. Focusing on laparoscopy we have extended distant light PS to handle short-range lights, developed methods for calibration, and an optimisation framework to achieve good solutions to depth. In contrast to other active methods tried in MIS, our approach can be used with an existing laparoscope with only minor modification. Unlike SfM, the approach handles textureless surfaces and does not require motion constraints. Unlike SfS, the method is stable and we can compute absolute depth. There is still further research to be done before it can handle unconstrained clinical conditions. Open challenges include handling spatially varying reflectance, handling time-varying illumination caused by changes in brightness or exposure, to learn different reflectance classes *a priori*, to handle tool occlusions, and to handle depth discontinuities with robust smoothing priors. We will also investigate how the recovered 3D can help solve the challenging problem of pre-operative/intra-operative registration.

References

1. Ackerman, J.D., Keller, K., Fuchs, H.: Surface Reconstruction of Abdominal Organs Using Laparoscopic Structured Light for Augmented Reality. In: 3DICA, pp. 39–46 (2002)
2. Brostow, G., Hernández, C., Vogiatzis, G., Stenger, B., Cipolla, R.: Video Normals from Colored Lights. PAMI, 2104–2114 (2011)
3. Burschka, D., Li, M., Ishii, M., Taylor, R.H., Hager, G.D.: Scale-invariant Registration of Monocular Endoscopic Images to CT-Scans for Sinus Surgery. MIA 9, 413–426 (2005)
4. Devernay, F., Mourgues, F., Coste-Manière, È.: Towards Endoscopic Augmented Reality for Robotically Assisted Minimally Invasive Cardiac Surgery. In: MIAR Workshop, pp. 16–20 (2001)
5. Grasa, Ó.G., Civera, J., Montiel, M.M.: EKF Monocular SLAM with Relocalization for Laparoscopic Sequences. In: ICRA, pp. 4816–4821 (2011)
6. Hu, M., Penney, G.P., Edwards, P.J., Figl, M., Hawkes, D.J.: 3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 68–77. Springer, Heidelberg (2007)
7. Hu, M., Penney, G.P., Rueckert, D., Edwards, P.J., Bello, F., Casula, R., Figl, M., Hawkes, D.J.: Non-rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 34–42. Springer, Heidelberg (2009)
8. Lau, W.W., Ramey, N.A., Corso, J.J., Thakor, N.V., Hager, G.D.: Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004, Part II. LNCS, vol. 3217, pp. 494–501. Springer, Heidelberg (2004)
9. Lim, J., Ho, J., Yang, M.H., Kriegman, D.: Passive Photometric Stereo from Motion. In: ICCV, pp. 1635–1642 (2005)

10. Malti, A., Bartoli, A., Collins, T.: Template-based Conformal Shape-from-Motion from Registered Laparoscopic images. In: MIUA (2011)
11. Mountney, P., Yang, G.-Z.: Motion Compensated SLAM for Image Guided Surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part II. LNCS, vol. 6362, pp. 496–504. Springer, Heidelberg (2010)
12. Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B., Hornegger, J.: Time-of-Flight 3-D Endoscopy. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 467–474. Springer, Heidelberg (2009)
13. Quartucci Forster, C.H., Tozzi, C.L.: Towards 3D Reconstruction of Endoscope Images Using Shape from Shading. GPI, 90–96 (2000)
14. Stoyanov, D., Darzi, A., Yang, G.Z.: Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004, Part II. LNCS, vol. 3217, pp. 41–48. Springer, Heidelberg (2004)
15. Stoyanov, D., Elson, D., Yang, G.Z.: Illumination Position Estimation for 3D Soft-tissue Reconstruction in Robotic MIS. In: IROS, pp. 2628–2633 (2009)
16. Woodham, R.: Photometric Method for Determining Surface Orientation from Multiple Images. Optical Engineering, 513–531 (1980)
17. Wu, C., Narasimhan, S.G., Jaramaz, B.: A Multi-image Shape-from-Shading Framework for Near-lighting Perspective Endoscopes. IJCV, 211–228 (2009)

Registration Accuracy: How Good Is Good Enough? A Statistical Power Calculation Incorporating Image Registration Uncertainty

Eli Gibson[1], Aaron Fenster[1,2,3,4], and Aaron D. Ward[2,4]

[1] Robarts Research Institute, London, Canada
[2] Lawson Health Research Institute, London, Canada
[3] Department of Oncology, The University of Western Ontario, London, Canada
[4] Department of Medical Biophysics,
The University of Western Ontario, London, Canada

Abstract. Image registration is an important tool for imaging validation studies investigating the effect of underlying focal disease on the imaging signal. The strength of the conclusions drawn from these analyses is limited by statistical power. Based on the observation that in this context, statistical power depends in part on uncertainty arising from registration error, we derive a power calculation formula relating registration error, sample size, and the minimum detectable difference between normal and pathologic regions on imaging. Statistical mappings between target registration error and fractional overlap metrics are also derived, and Monte Carlo simulations are used to evaluate the derived models and test the strength of their assumptions.

Keywords: imaging validation, registration error, statistical power.

1 Introduction

Registration of medical images can enable complex analyses of medical data as well as image-guided diagnosis and treatment, provided the registration is performed with sufficient accuracy. There can be tradeoffs associated with achieving higher accuracy [1], including greater human interaction to guide registration algorithms to correct solutions, higher required image quality, and higher computational cost. Thus, for each study, it is important to identify the maximum acceptable level of registration error. This threshold is application-dependent [2], and establishing application-specific thresholds for maximum acceptable error has been identified as a key challenge in the field [1,2].

In image-guided interventions (IGI), registration can be used to guide a tool tip to a target region. Studies of such systems usually involve quantifying either the distance from the tool tip to the target or the overlap of the tool's treatment volume with the target. In some IGI applications (e.g. aortic aneurysms [3] and prostate cancer biopsy [4]), acceptable registration error thresholds have been identified for specific anatomy and imaging modalities. However, in some IGI

contexts (e.g. keyhole brain surgery for ablation of epileptogenic foci [5]), it is currently unclear how to localize the targets on preoperative imaging. This motivates the need for studies to address this question.

In studies of the utility of medical imaging for disease localization (henceforth, *imaging validation studies*), registration can be used to align images of the medical imaging modality to be studied (*study images*) with *reference images* wherein ground truth regarding localized disease is defined (e.g. on pathologist-contoured digital pathology images). Such studies involve the measurement of the effect of the presence of disease *features of interest* (e.g. cancerous tissue) on image intensity (or other derived quantities) on the corresponding region of interest on the study image. Each ground truth delineation of disease features of interest on the reference image is mapped by an ideal (0 error) registration to a region of interest on the study image (denoted as R hereafter). We denote as $\tilde{R}$ such a region determined by a non-ideal (> 0 error) registration. Thus, in contrast to the IGI scenario, for imaging validation studies, the fidelity of the $R - \tilde{R}$ *regional overlap* is paramount. Mapping errors that result in smaller overlap may lead to larger required sample sizes to achieve a given minimum detectable difference (MDD) on imaging between pathologic and benign regions. This observation leads to three key questions affecting study design. **(1) What is the maximum acceptable registration error?** For a fixed sample size and desired MDD, what is the maximum acceptable image registration error? **(2) How many subjects are needed?** For a known image registration error and desired MDD, what is the required sample size? **(3) What is the minimum detectable difference?** For a fixed sample size and known image registration error, what is the MDD? To the best of our knowledge, there has been no previous work in the literature addressing these questions in the context of imaging validation studies based on image registration.

As a first step toward fully answering these questions, in this paper, we provide a derivation that yields the relationship between image registration error, sample size, and MDD, where image registration is used to determine whether the presence of particular anatomy, pathology or other features of interest in the underlying tissue is reflected in a change in the mean intensity of study image voxels containing the features of interest. The derivation of a statistical power calculation that incorporates uncertainty due to registration error yields a set of three equations that can be used to answer the three questions enumerated above. Statistical power is a measure that describes the probability of a study finding a statistically significant result when there is an underlying difference to be found. Thus, for studies to determine whether focal disease affects study image intensity, the acceptable registration error is defined relative to the study's statistical power. The statistical power is a relationship between the size of the study, the acceptable levels of type I error (false positive results from the study) and type II error (false negative results from the study), the intensity distributions in R and in the background, and the registration error. This statistical power is commonly expressed in the form of a sample size calculation that relates how many subjects to recruit for a particular study design or an MDD

calculation that relates how small a difference can be detected for a particular study design. To the best of our knowledge, this work represents the first derivation of a statistical power calculation for medical imaging validation studies that incorporates uncertainty in the overlap of R and $\tilde{R}$ due to registration error.

The remainder of this paper outlines the derivation of the relationship between registration error and statistical power for one study design (Section 2), describes simulations used to validate components of the derivation (Section 3), presents the results of these simulations (Section 4) and discusses the implications of these relationships (Section 5).

2   Derivation of MDD and Sample Size Relative to Registration Error

The statistical power calculation depends in part on the type of statistical analysis used in the study. In this paper, we addressed a specific analysis that statistically compares a pool of samples drawn from multiple identified $\tilde{R}$ to another pool drawn from background regions B using a t-test on the sample intensities. For imaging validation studies, image registration error is preferably measured as the target registration error (TRE), since a post-registration comparison of accurate segmentations is not feasible due to the absence of knowledge of the boundary of R on the study image. To derive the relationship between the TRE and inferential power, we utilize an intermediate metric, fractional overlap (FO), which is the ratio of the volume of the intersection of R and $\tilde{R}$ to the volume of $\tilde{R}$. The following two sections will derive the relationships between the MDD, sample size and the FO, and the relationship between TRE and FO.

2.1   Mapping Registration Error to Fractional Overlap

Fractional overlap ($\frac{|\tilde{R} \cap R|}{|\tilde{R}|}$) of two registered spherical regions R and $\tilde{R}$ can be expressed as a function of the radius of the regions r and the 3D registration error $\boldsymbol{x}$ between their centers: $f = \frac{\pi(4r+ry)(2r-ry)^2/12}{4\pi r^3/3} = (y^3 - 12y + 16)/16$ for $y \leq 2$ otherwise $f = 0$, where $y = ||\boldsymbol{x}||/r$ is the relative error.

The probability density function (PDF) F of FO can be derived as a function of the PDF Y of the relative error under certain assumptions: (1) each R and $\tilde{R}$ are spherical and of a fixed size, and (2) the registration error can be modeled as a 3D Gaussian. For $f = 0$, $p(F = 0) = p(Y >= 2)$. For $f > 0$, we use the relation for functions of random variables, $p(F = f) = p(Y = y(f))/|f'(y(f))|$. The derivative $\frac{df}{dy} = (3y^2 - 12)/16$. As $(y^3 - 12y + 16)/16 - f = 0$ has 3 real roots for $0 \leq f \leq 2$, we can express the inverse using the trigonometric expressions for cubic roots. There is one solution in the range $0 \leq y \leq 2$: $f^{-1}(y) = y(f) = 4cos(\frac{acos(1-f)+\pi}{3})$. Combining these intervals,

$$p(F = f) = \delta(f)p(Y >= 2) + \frac{16p(Y = 4cos(\frac{acos(1-f)+\pi}{3}))}{|3(4cos(\frac{acos(1-f)+\pi}{3}))^2 - 12|}. \tag{1}$$

For a registration error vector $\mathbf{x}$ that is distributed as a 3D Gaussian with component-wise variance a^2, (i.e. $\mathbf{X} \sim N_3(0, a^2 I)$), the registration error $x = ||\mathbf{x}||$ has a Maxwell-Boltzmann distribution. By a change of variables, $p(Y = y) = rp(X = yr)$, and, $p(Y > 2) = p(X > 2r)$, which can be substituted into Equation 1, yielding the FO as a function of registration error

$$p(F = f) = \delta(f)(1 - (erf(\frac{2r}{\sqrt{2}a}) - \sqrt{\frac{2}{\pi}} \frac{2r exp(-\frac{(2r)^2}{2a^2})}{a})) +$$

$$\frac{16r\sqrt{\frac{2}{\pi}}(4cos(\frac{acos(1-f)+\pi}{3})r)^2 exp(-(4cos(\frac{acos(1-f)+\pi}{3})r)^2/(2a^2))/a^3}{|3(4cos(\frac{acos(1-f)+\pi}{3}))^2 - 12|}. \qquad (2)$$

For FO, the mean $\mu_F(\frac{a}{r}) = \int_0^1 fp(F = f)df$ and standard deviation $\sigma_F^2(\frac{a}{r}) = \int_0^1 (f - \mu_F(\frac{a}{r}))^2 p(F = f)df$ vary with the ratio of the TRE scaling parameter to the radius of R, and are invariant to specific choices of a and r. Integrating numerically yields the relationships shown in Figure 1.
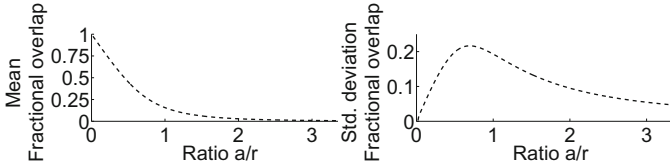
Fig. 1. Mean (left) and std. (right) fractional overlap as a function of the ratio a/r of the target registration error scaling factor to the radius of R.

2.2 Relationship between MDD, Sample Size and Fractional Overlap

The derivation is made under assumptions that (1) intensities of voxels containing the *features of interest* and background are independently distributed as $I_R \sim N(\mu_R, \sigma_R^2)$ and $I_B \sim N(\mu_B, \sigma_B^2)$, respectively; (2) statistical analysis will be performed by an unpaired two-sample heteroscedastic T-test of the null hypothesis that $\mu_R = \mu_B$ against the alternative hypothesis that $\mu_R \neq \mu_B$; (3) the number of measurements from each $\tilde{R}$ is constant across samples; (4) the number n of regions $\tilde{R}$ is large enough that the mean FO approximates a normal distribution (by the central limit theorem); and (5) the number of voxels v in each $\tilde{R}$ is large enough that discretizing error can be ignored.

When there is no registration error, the minimal detectable difference μ_d between μ_R and μ_B using a two sample t-test can be expressed as $\mu_d = T\sqrt{(\sigma_R^2 + \sigma_B^2)/(nv)}$, where T is a statistical threshold $t_{\alpha\{2\},nv} + t_{\beta\{1\},nv}$, where $t_{\alpha\{2\},nv}$ and $t_{\beta\{1\},nv}$ are taken from two- and one-tailed t-tables with nv degrees of freedom, constraining type I error to α and type II error to β.

When there is misregistration of the i-th region, the measurements in $\tilde{R}$ may contain samples from the background. Given FO f_i, the sample mean is

$\sum_i (\sum_{j=1}^{f_i v} (I_{j,i,R}) + \sum_{j=1}^{(1-f_i)v} (I_{j,i,B}))/(vn)$. As each $I_{j,i,R}$ and $I_{j,i,B}$ is a Gaussian random variable, the distribution of the mean is:

$$N(\sum_i (f_i \mu_{i,R} + (1 - f_i)\mu_{i,B})/n, (\sum_i (f_i \sigma_{i,R}^2 + (1 - f_i)\sigma_{i,B}^2)/(n^2 v)), \quad (3)$$

or, by substituting $\mu_d = \mu_R - \mu_B$ and $\sigma_d^2 = \sigma_R^2 - \sigma_B^2$,

$$N(\mu_d \sum_i (f_i)/n + \mu_B, (\sigma_d^2 \sum_i (f_i)/n + \sigma_B^2)/(nv)). \quad (4)$$

Because the FOs f_i are random variables contributing to both the mean and standard deviation of the distribution, the mean distribution is not Gaussian. To simplify the model, we introduce two approximations. First, we approximate $\sum_{i=1}^{n} (f_i)/n$ with a random variable $\sim N(\mu_F, \sigma_F^2/n)$ in the mean, using the central limit theorem approximation for sufficiently high n. Second, we approximate $\sum_i^n (f_i)/n$ as μ_F in the standard deviation. The resulting distribution $N(\mu_d N(\mu_F, \sigma_F^2/n) + \mu_B, (\mu_F \sigma_d^2 + \sigma_B^2)/(nv))$ can be simplified to $N(\mu_d \mu_F + \mu_B, (\mu_F \sigma_d^2 + \sigma_B^2)/(nv) + \mu_d^2 \sigma_F^2/n)$.

Because this model for the distribution of the mean is Gaussian, as in the errorless case, we can incorporate this into the normal power analysis framework by constructing a hypothetical population that would have the same mean distribution: $N(\mu_d \mu_F + \mu_B, \mu_F \sigma_d^2 + \sigma_B^2 + \mu_d^2 \sigma_F^2 v)$. The pooled variance for this analysis will be $(\sigma_B^2 + \mu_F \sigma_d^2 + \sigma_B^2 + \mu_d^2 \sigma_F^2 v)/2$ or, simplified, $\mu_F \sigma_d^2/2 + \sigma_B^2 + \sigma_F^2 \mu_d^2 v/2$. The MDD between the $\tilde{R}$ and background can be expressed in terms of μ_d as

$$\mu_d \mu_F + \mu_B - \mu_B = \sqrt{\frac{\mu_F \sigma_d^2 + 2\sigma_B^2 + \sigma_F^2 \mu_d^2 v}{nv}} T. \quad (5)$$

Solving for mean FO μ_F yields

$$\mu_F = \frac{\sigma_d^2 T^2 \pm T \sqrt{\sigma_d^4 T^2 + 8n\sigma_B^2 \mu_d^2 v + 4nv^2 \mu_d^4 \sigma_F^2}}{2\mu_d^2 nv}. \quad (6)$$

Solving for the sample size yields

$$n = T^2 \left(\frac{2\sigma_B^2 + \mu_F \sigma_d^2}{\mu_d^2 v \mu_F^2} + \frac{\sigma_F^2}{\mu_F^2} \right). \quad (7)$$

Solving for the MDD yields

$$\mu_d = \sqrt{\frac{2\sigma_B^2 + \mu_F \sigma_d^2}{nv(\mu_F^2 - T^2 \sigma_F^2/n)}} T. \quad (8)$$

3   Simulations

We performed Monte Carlo simulations to assess the accuracy of the derived statistical model, and the sensitivity of the model to assumption violations.

Table 1. Power simulation parameters. Values were specified as [a,b,c], ranges as [a..b], and values sampled once per $\tilde{R}$ from a normal distribution as N(mean, std.).

	$I_R - I_B$	n	v	a/r	α	β	σ_R^2	σ_B^2
S1	MDD	30	30	[0.1,0.5,1]	0.05	0.8	[1..100]	[1..100]
S2	MDD	[1..30]	30	1	0.05	0.8	10	10
S3	MDD	30	[1..30]	1	0.05	0.8	10	10
S4	MDD	30	N(30,[0..10])	1	0.05	0.8	10	10

To assess the model relating MDD, sample size and FO, we sampled N sets of image intensities from the background and $\tilde{R}$ intensity distributions and performed two sample T-tests of the null hypothesis that sample mean intensities were equal. In each simulation, $\mu_R - \mu_B$ was set to the MDD predicted by the model, and N=160,000 samples were taken (to yield a 95% confidence interval of width 0.5% on β). The proportion of positive t-test results from the simulation should match the model's type II error of $1 - \beta$. We assessed (S1) the accuracy of the model under the assumptions, as well as the sensitivity of the model to violations of the assumptions regarding (S2) the number of regions $\tilde{R}$ sampled, (S3) the number of voxels per $\tilde{R}$, and (S4) the constancy of the $\tilde{R}$ volume. The parameters varied in these simulations are described in Table 1.

To assess the model relating FO to registration error, we sampled error vectors $\boldsymbol{x}$ from a 3D Gaussian distribution and calculated the FO of R and $\tilde{R}$ with centers offset by $\boldsymbol{x}$. The resulting empirical PDF was compared to the PDF predicted by our model. We assessed (S5) the accuracy of the model under the given assumptions, with a ranging from $\frac{2}{100}$ to $\frac{350}{100}$ and $r = 10$.

4   Results

Simulation results from S1 through S4 are summarized in Fig. 2(a-d). The y-axes indicate the difference between the power predicted by the model and the simulations. A value of 0% indicates that the model exactly predicted the simulation results. Values of $-x\%$ and $+x\%$ indicate that the model under- and overestimated the power, respectively. Fig. 2(a) shows that when registration error is large (i.e. high a/r), the model underestimates power, particularly with small sample sizes. Fig. 2(b) shows that for small sample sizes, the model underestimates power, particularly with large registration errors. Fig. 2(c) shows that the model's estimate of power is reliable except in cases where small numbers (< 5) of point samples (e.g. voxels) are taken from each $\tilde{R}$. Fig. 2(d) shows that the model's estimate of power is accurate and robust to variance in the number of point samples taken from each $\tilde{R}$. In simulation S5, the model predicted the simulated mean and std. FO as a function of a/r (Fig. 1) to within 0.0006.
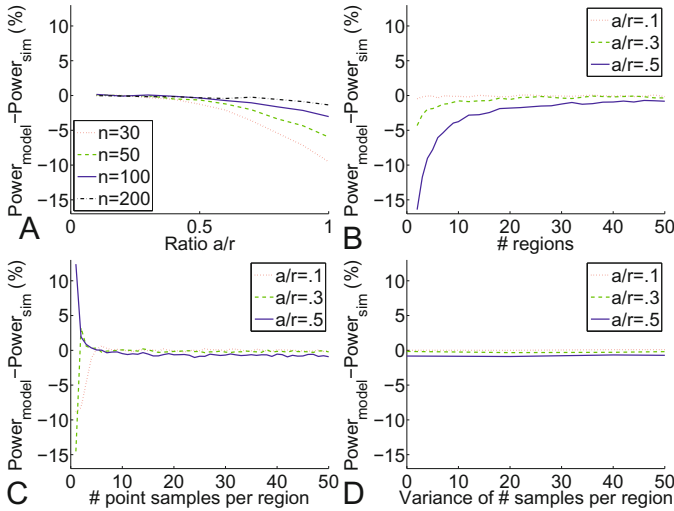
Fig. 2. (a) S1: Power vs. registration error a/r for several sample sizes. (b) S2: Power vs. sample size. (c) S3: Power vs. number of samples (e.g. voxels) / $\tilde{R}$. (d) S4: Power vs. variance in number of samples (e.g. voxels) / $\tilde{R}$.

5 Discussion

This work provides a derivation of a statistical power calculation incorporating image registration uncertainty and addressing three central questions in the design of imaging validation studies. **(1) Eq. (6):** What is the maximum acceptable registration error? **(2) Eq. (7):** How many subjects are needed? **(3) Eq. (8):** What is the MDD between normal and pathologic image regions? We derived the relationship between the scaling parameter of a 3D Gaussian TRE and the distribution of FO of spherical tumours. We also derived an approximate relationship between an arbitrary distribution of FO and the statistics of a study design. The combination of these derivations elucidated a relationship between registration error, sample size and statistical power, yielding a set of three equations that are central to the design of imaging validation studies.

These relationships could be used in several applications. During study design, Eq. (7) or (8) could be used to evaluate or control the power, after estimating imaging properties and registration errors, while Eq. (6) could be used to guide the selection of registration algorithms under the constraint of a study design. During algorithm development, Eq. (6) could be used to assess whether an algorithm has sufficient accuracy for a particular application.

We ran Monte Carlo simulations to test the fidelity of our model both when our assumptions were met and when some of them were relaxed. Our results showed that (1) the model predicts statistical power reliably for reasonable registration error (i.e. not larger than 50% of the radius of R) and the sample size > 30 (Fig. 2(a-b)); (2) the model predicts power reliably when > 5 samples (e.g. voxels) are obtained from within each $\tilde{R}$ (Fig. 2(c)); (3) the model predicts

power reliably regardless of the variability in the number of samples obtained from within each $\tilde{R}$ (Fig. 2(d)); and (4) the model accurately predicted the FO as a function of registration error.

The limitations of this work lie mainly in the strong assumptions made by the derivations. Although we have tested the robustness of the model to the relaxation of some of these assumptions, our testing in this regard is not exhaustive. Furthermore, extensions of these models may allow some assumptions to be relaxed (e.g., assumptions of spherical regions, isotropic 3D Gaussian registration error, and no correlation of voxels within each R). Also, our derivation is based on a relatively simple (albeit useful) statistical design; because analysis of statistical power depends on the statistical designs used, it would be valuable to extend the presented derivations to account for paired tests (to account for voxel intensity correlations within subjects), cluster randomization (to allow for intensity correlations within each R), regression (for longitudinal analyses) and multivariate data.

References

1. Simon, D., O'Toole, R.V., Blackwell, M., Morgan, F., Digioia, A.M., Kanade, T.: Accuracy validation in image-guided orthopaedic surgery. In: Proc. Int. Symp. Medical Robotics and Computer Assisted Surgery, pp. 185–192 (1995)
2. Loew, M.H., Rodriguez-Carranza, C.E.: Technical issues in multimodality medical image registration. In: Proc. IEEE Symp. Computer-Based Medical Systems, pp. 2–7 (1998)
3. Penney, G., Varnavas, A., Dastur, N., Carrell, T.: An Image-Guided Surgery System to Aid Endovascular Treatment of Complex Aortic Aneurysms: Description and Initial Clinical Experience. In: Taylor, R.H., Yang, G.-Z. (eds.) IPCAI 2011. LNCS, vol. 6689, pp. 13–24. Springer, Heidelberg (2011)
4. van de Ven, W.J.M., Litjens, G.J.S., Barentsz, J.O., Hambrock, T., Huisman, H.J.: Required Accuracy of MR-US Registration for Prostate Biopsies. In: Madabhushi, A., Dowling, J., Huisman, H.J., Barratt, D. (eds.) Prostate Cancer Imaging 2011. LNCS, vol. 6963, pp. 92–99. Springer, Heidelberg (2011)
5. Eriksson, S.H., Free, S.L., Thom, M., Harkness, W., Sisodiya, S.M., Duncan, J.S.: Reliable registration of preoperative MRI with histopathology after temporal lobe resections. Epilepsia 46(10), 1646–1653 (2005)

Joint Tumor Segmentation and Dense Deformable Registration of Brain MR Images[⋆]

Sarah Parisot[1,2,3], Hugues Duffau[4],
Stéphane Chemouny[3], and Nikos Paragios[1,2]

[1] Center for Visual Computing, Ecole Centrale Paris, Chatenay Malabry, France
[2] Equipe GALEN, INRIA Saclay - Ile de France, Orsay, France
[3] Intrasense SAS, Montpellier, France
[4] Département de Neurochirurgie, Hopital Gui de Chauliac, CHU Montpellier, France

Abstract. In this paper we propose a novel graph-based concurrent registration and segmentation framework. Registration is modeled with a pairwise graphical model formulation that is modular with respect to the data and regularization term. Segmentation is addressed by adopting a similar graphical model, using image-based classification techniques while producing a smooth solution. The two problems are coupled via a relaxation of the registration criterion in the presence of tumors as well as a segmentation through a registration term aiming the separation between healthy and diseased tissues. Efficient linear programming is used to solve both problems simultaneously. State of the art results demonstrate the potential of our method on a large and challenging low-grade glioma data set.

1 Introduction

The automatic evaluation of the evolution of brain tumors, that are often monitored through MRI imaging, is of growing interest. The methods currently used to evaluate precisely the position, size and evolution of a tumor, involve a complete 3D manual segmentation by an expert. It is, however, extremely time consuming and highly dependent on the expert, particularly in the case of the diffusively infiltrative low grade gliomas, that have fuzzy boundaries and inhomogeneous appearances [1]. As it has been shown that low-grade gliomas tend to appear in preferential locations [2], the construction of statistical atlases enables the study of location dependent behaviour of tumors. As a result, both automatic segmentation and registration with a healthy atlas of tumor-bearing images are of great interest for the study of brain tumors.

State of the art segmentation methods combine efficient classification techniques [3] with low level segmentation methods [4]. From such perspective, tumor detection is addressed as a classification problem where one aims at separating healthy from diseased tissues at the voxel level, while imposing smoothness

---

constraints. Despite promising performance, these methods suffer from modularity and scalability. Visual and geometric appearance of tumors depend on organs, contrast agent, hardware acquisition device. The use of segmentation-by-deformable registration of anatomical atlases [5] can be an answer to the aforementioned limitations due to their ability to include global consistency and deal with isolated erroneous measurements. This kind of method however suffers from the inherent link between the two problems, making the segmentation dependent on the registration quality and vice versa. A common approach to deal with the presence of tumors for registration is to mask the pathology after segmentation [6]. Another strategy is atlas seeding and simulating the tumor induced deformation in the atlas using growth models [7].

The aim of this paper is to perform tumor segmentation and registration with missing correspondences in a one shot optimization approach. Concurrent segmentation and registration is a not a new objective [8,9], however the problem becomes far more challenging in the presence of tumors. This is due to the fact that modeling tumor appearance and geometry is not as trivial as modeling anatomical regions, while at the same time registration is violated in the tumor area. [10] proposes registration of an atlas, deformed using a complex tumor growth model, to the patient's space. The parameters are learned using the Expectation Maximization algorithm which can be very computationally expensive and sensitive to initialization. Our approach aims at combining the modeling of both problems where the unknown variables correspond to a two layer graphical model, one that represents the 3D deformation field and another that refers to the 3D binary segmentation map. This graphical model is superimposed to the volume domain. The deformation is obtained using the discrete formulation introduced in[11]. Segmentation is obtained through a conventional graph-based formulation [12] where tumors and healthy tissues are separated through the use of boosting. In order to reduce complexity we adopt a reduced label grid that estimates segmentation likelihoods in the whole image domain through interpolation of the associated variables. The two layers are interconnected with a combined cost that relaxes the registration in the presence of tumors, while at the same time performs a segmentation-by-registration using the segmentation costs as criterion. Linear programming and duality [13] are used to determine the optimal solution of the combined problem. The proposed formulation is modular with respect to the registration criterion and the segmentation likelihoods, scalable since the segmentation/registration grid can be adapted to the application domain and computationally efficient. The remainder of this paper is organized as follows, section 2 describes the combined registration and segmentation method, while experimental validation is presented in the following section. Discussion and future directions conclude the paper.

2   Graph-Based Image Registration and Segmentation

Let us now first introduce the individual components of the method and then conclude the methodological part of the paper with their coupling. The resulting

formulation refers to a two-layer interconnected graph where both the segmentation labels on the atlas space and the deformable registration field are determined through a one shot optimization. Such a formulation seeks both maximization of the posterior statistical likelihoods of the tumor versus healthy voxels, and optimal similarity matching between the atlas and the observed volume that is relaxed in the tumor area.

2.1 Graph-Based Image Registration

In the task of image registration, we seek to align a source image A to a target image I defined on a domain $\Omega \subset \mathbb{R}^3$. To this end, we seek to estimate a geometric deformation $\mathcal{T}(\mathbf{x})$ that will map A to I.

Let us consider a sparse grid $\mathcal{G} \subset \Omega$ superimposed to the image. The deformation is evaluated on the grid's control points $\mathbf{p}$ as $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \sum_{p \in \mathcal{G}} \eta\left(\|\mathbf{x} - \mathbf{p}\|\right) \mathbf{d_p}$, where $\mathbf{d_p}$ is the displacement vector of the control point $\mathbf{p}$. The idea is to deform the source image by moving the grid's control points. The displacement of a voxel $\mathbf{x}$ of the image will be determined by the control point's displacements and the influence of each control point on $\mathbf{x}$, which is given by the projection function η. A typical example of projection function would be cubic B-splines. The optimal deformation field should minimize the distance between the target image $I(\mathbf{x})$ and deformed image $A(\mathcal{T}(\mathbf{x}))$, that is evaluated by a similarity function ρ:

$$E(\mathcal{T}) = \frac{1}{|\mathcal{G}|} \sum_{\mathbf{p} \in \mathcal{G}} \int_{\Omega} \bar{\eta}\left(\|\mathbf{x} - p\|\right) \rho\left(I(x), A(\mathcal{T}(\mathbf{x}))\right) dx \tag{1}$$

The similarity function being defined on the image domain Ω, a function $\bar{\eta}$ is introduced to back project the voxel wise information on the grid's control points. It determines how much voxel $\mathbf{x}$ influences the control point $\mathbf{p}$.

We adopt the discrete approach described in [11], in which problem (1) is reformulated as a labeling problem. To this end, we define a set of labels $\mathcal{L} = \{l^1, ..., l^k\}$ and a discrete displacement set $\Delta = \{\mathbf{d^1}, ..., \mathbf{d^k}\}$. Each label corresponds to a specific displacement vector so that assigning a label $l_p \in \mathcal{L}$ to a control point $\mathbf{p}$ imposes the corresponding displacement $\mathbf{d^i}$ of the node. The deformation field can be reformulated as: $\mathcal{T}_l(\mathbf{x}) = \mathbf{x} + \sum_{p \in \mathcal{G}} \eta\left(\|\mathbf{x} - \mathbf{p}\|\right) \mathbf{d^{l_p}}$. This enables to pose the problem as a discrete Markov Random Field (MRF) optimization with respect to the labeling l, where the goal is to minimize the following energy:

$$E_{def}(l) = \frac{1}{|\mathcal{G}|} \sum_{p \in \mathcal{G}} V_p(l_p) + \lambda \sum_{p \in \mathcal{G}} \sum_{q \in \mathcal{N}(p)} V_{pq}(l_p, l_q) \tag{2}$$

$V_p(l_p)$ can be approximated as $V_p(l_p) \approx \int_{\Omega} \bar{\eta}\left(\|\mathbf{x} - \mathbf{p}\|\right) \rho\left(I(x), A\left(\mathbf{x} + \mathbf{d^{l_p}}\right)\right) dx$, V_{pq} is a pairwise constraint that imposes local smoothness of the deformation field and λ is a constant parameter balancing the single and pairwise costs. This formulation is not sufficient in the case of lacking correspondences between the images. In these areas, the source image will be heavily deformed and information will be lost.

2.2   Graph-Based Image Segmentation

In order to automatically segment tumor voxels, we seek to construct a classi-
fier that will describe the likelihood that each voxel belongs to a tumor. One
simple and efficient way to do so is to learn the classifier based on a vector of
image features and boosting. We use the Gentle Adaboost algorithm [3], that is
known to be more robust to noise and outliers than other boosting algorithms.
We adopt three feature spaces. First, we rely on the intensity values in the voxel
of interest $\mathbf{x_i}$ and its neighborhood. We include in the feature vector an intensity
patch centered on $\mathbf{x_i}$, the median, entropy and standard deviation of patches of
variable sizes. Second, we use Gabor features [14] on 2 scales and 10 orientations.
Last, the tumor's presence will introduce some lack of symmetry between the
hemispheres. Assuming we know an approximate symmetry plane Π, we com-
pute a symmetry feature as $Sym(\mathbf{x_i}) = \frac{1}{k}\sum_{\mathcal{N}(\mathbf{x_i})} I(\mathbf{x_i}) - \frac{1}{k}\sum_{\mathcal{N}(\mathbf{x_i})} I(\mathbf{x_i}_\Pi)$, $\mathbf{x_i}_\Pi$
being the symmetric of $\mathbf{x_i}$ with respect to Π, and $\mathcal{N}$ a neighborhood around
$\mathbf{x_i}$ introduced to reduce the approximation error. The classifier's output can
be converted into a tumor $p_{tm}(\mathbf{x})$ and a background probability $p_{bg}(\mathbf{x})$. Both
probabilities are fixed after training.

The problem of tumor segmentation can be cast as an MRF optimization on
the whole volume I, where each voxel gets a label $l \in \{0, 1\}$ identifying it as
tumor or background:

$$E_{seg}(l) = \sum_{\mathbf{x}} V_x(l(\mathbf{x})) + \lambda_{seg} \sum_{\mathbf{x}} \sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} V_{xy}(l(\mathbf{x}), l(\mathbf{y})) \tag{3}$$

The single costs impose a labeling according to the classification likelihoods:
$V_x(l(\mathbf{x})) = -\log(p_{bg}(\mathbf{x}))(1-l((\mathbf{x}))) - \log(p_{tm}(\mathbf{x}))\, l((\mathbf{x}))$. V_{xy} is a pairwise cost
added to impose local consistency of the segmentation and λ_{seg} balances the
single and pairwise costs. The segmentation task could also benefit from the
registration, since existing correspondences between the atlas and the image to
be segmented could eliminate false detections of the classifier.

2.3   Combined Registration and Segmentation

Let us now describe the main contribution of this paper: the joint segmentation-
registration framework. Let us consider that the source image A, a healthy brain
image, is to be registered to a target brain image I featuring a low-grade glioma.
In the method presented in section 2.1, the tumor's presence leads to heavy
deformation of the brain and loss of structural information. The aim of our
method is to amend this problem by detecting the tumor's position and not
taking it into account during registration, while at the same time obtaining a
precise segmentation of the tumor.

We seek to evaluate the geometric deformation $\mathcal{T}(\mathbf{x})$ and the tumor's position
$\mathcal{S}(\mathbf{x})$ through the same optimization. We adopt a discrete MRF formulation in
which each grid node $\mathbf{p}$ is assigned a label $l \in \mathcal{L} = \{l^1, ..., l^{2k}\}$. Each label cor-
responds to a pair $\{s, \mathbf{d}\} \in \{0, 1\} \times \{\mathbf{d^1}, ..., \mathbf{d^k}\}$. The first term s^{l_p} will simply

characterize the node as tumor or background, while the second one $\mathbf{d^{l_p}}$ determines the node's displacement. The segmentation and displacement information is propagated to the whole image by cubic B-spline interpolation:

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \sum_{p \in \mathcal{G}} \eta(\|\mathbf{x} - p\|)\mathbf{d^{l_p}} \quad \text{and} \quad \mathcal{S}(\mathbf{x}) = H(\sum_{p \in \mathcal{G}} \eta(\|\mathbf{x} - p\|)s^{l_p} - 0.5) \quad (4)$$

Due to the interpolation, the segmentation map $\mathcal{S}(\mathbf{x})$ is not binary. We deal with this issue by thresholding the map using the heaviside function H.

Let us now proceed with the definition of the singleton cost $V_p(l_p)$. In the case of healthy voxels, the evaluation of the deformation is still based on a similarity measure between the source and target image. In the tumor area ($s^{l_p} = 1$), we cannot rely on the similarity measure and therefore introduce a constant cost C_{tm} as suggested in [15] for stereo matching:

$$V_{def}(l_p) = \int_\Omega \bar{\eta}\left(\|\mathbf{x} - \mathbf{p}\|\right)\left((1 - s^{l_p})\,\rho\big(V(\mathbf{x}), A(\mathbf{x} + \mathbf{d^{l_p}})\big) + s^{l_p}C_{tm}\right) dx \quad (5)$$

This potential enables registration without taking into account the area detected as tumor (highly dissimilar points in the image). While part of the tumor can be detected using this single potential, the lack of similarity alone is not sufficient to efficiently segment the tumor. Only tumors voxels with high contrast enhancement will be detected. Furthermore, dissimilarity doesn't necessarily correspond to a tumor. To enhance the segmentation and prevent false positives, we make use of the classifier built in section 2.1. We couple the deformation potential with a second one acting mainly on the segmentation space:

$$V_{seg}(l_p) = \int_\Omega \bar{\eta}\left(\|\mathbf{x} - \mathbf{p}\|\right)\left(-\log\big(p_{bg}(\mathbf{x} + \mathbf{d^{l_p}})\big)\big(1 - s^{l_p}\big) - \log\big(p_{tm}(\mathbf{x} + \mathbf{d^{l_p}})\big)\,s^{l_p}\right) dx \tag{6}$$

This potential imposes that voxels with a high probability of being tumor (p_{tm}) are labeled accordingly. It takes the deformation into account since the classifier's score is defined on the target image: the better the source image is aligned to the target image, the more precise becomes the tumor's position. We can now define the singleton cost as $V_p(l_p) \approx \alpha V_{def}(l_p) + (1 - \alpha)V_{seg}(l_p)$, where α is a constant coefficient of key importance balancing the 2 potentials' influence. If α is high, the registration will be prioritized, resulting in a good registration (except in the tumor area) and poor segmentation. On the contrary, imposing a low α yields a good segmentation (assuming the classifier is efficient), while the registration quality will be deteriorated. As for the pairwise cost, it is set as:

$$V_{pq}(l_p, l_q) = \alpha\lambda\frac{\|\mathbf{d^{l_p}} - \mathbf{d^{l_q}}\|}{\|\mathbf{p} - \mathbf{q}\|} + (1 - \alpha)\frac{|s^{l_p} - s^{l_q}|}{\|\mathbf{p} - \mathbf{q}\|} \quad (7)$$

In order to recover the optimal labeling solution, we use Fast-PD[13], an efficient MRF optimization method based on linear programming.

3   Experimental Validation

We adopt a multi-scale incremental approach (both on the image and grid reso-
lutions) were the tumor presence will have an increasing impact on the process.
We evaluated the deformation and segmentation on 3 image levels and 4 grid
levels ($9 \times 9 \times 5$ to $65 \times 65 \times 37$). The parameter C_{tm} was increased at each
level, from 5 to 10 times the mean value of the similarity measure, while the
parameter α is progressively reduced from 1 to 0.015. λ and λ_{seg} were set at 20
and 3 respectively. At each level t, new deformation and segmentation fields are
estimated w.r.t. the source image deformed in level $t-1$. The deformation field
$\mathcal{T}^t(\mathbf{x})$ is updated by composition with the one obtained at level $t-1$. Since false
positives are less likely to appear in the coarse levels, we add a penalty term as
$V^t_{pen}(\mathbf{l_p}) = -\int_\Omega \bar{\eta}\left(\|\mathbf{x}-\mathbf{p}\|\right)(1-s^{l_p})\exp(-t)dx$ to voxels labeled as background
($S^{t-1}(\mathbf{x}) = 0$). We use the Sum of Absolute Differences as similarity criterion.

The healthy brain template used for registration was a 3D MRI FLAIR image,
of size $256 \times 256 \times 24$, and resolution $0.9 \times 0.9 \times 5.45mm^3$. All images in our
data-set (93 images) were 3D FLAIR images of different patients with low-grade
gliomas that have been manually segmented by an expert. Their sizes ranged
from $256 \times 256 \times 24$ to $512 \times 512 \times 33$, and resolution from 0.4×0.4 to.$9 \times 0.9mm^2$
in the (x,y) plane and 5.3 to 6.4 mm in the z plane. The smallest tumor was
$3.5\ cm^3$, and the largest was $230cm^3$. Preprocessing involved skull stripping and
intensity regularization. In order to avoid losing tumor vs healthy tissue contrast,
we simply set all images to the same median and interquartile range as the
reference pose, without taking into account background voxels. Eventually, all
images were rigidly registered to the reference pose, which permitted to use the
template's symmetry plane to evaluate the symmetry feature for the boosting
classifier. 36 randomly selected volumes were used for boosting learning. The
patches sizes for intensity statistics were $k \times k \times 3$, with $k = \{3, 5, 7\}$.

We tested our framework on the 57 remaining images. Overall computational
time was approximately 6 minutes. The registration was compared qualitatively
to individual registration where the pathology is masked using the manual seg-
mentation masks (enabling comparison to a sequential approach). Some visual
registration results, along with the corresponding deformation maps are shown
in Fig. 2. We observe high visual correspondences between the target and reg-
istered images outside the tumor area for both methods. As shown by the indi-
vidual registration's deformation maps, deformations inside the tumor area are
very important and excessively unnatural. Using our method, the deformations
are much smoother and the deformed image's anatomy is respected, showing
that our method performs significantly better in the tumor area. The automatic
segmentation (A) was evaluated w.r.t the manual segmentation (M) using the
Dice coefficient, the rate of false positives ($FP = \frac{\|A\|-\|M\cap A\|}{\|A\|}$) and the Mean
Absolute Distance (MAD) between contours. Results were compared to the voxel
wise individual segmentations. Comparative boxplots are shown in Fig. 1. The
median dice increases from 77% to 80%, while false positives significantly dimin-
ished (median 30% to 20%) and MAD values also diminished.
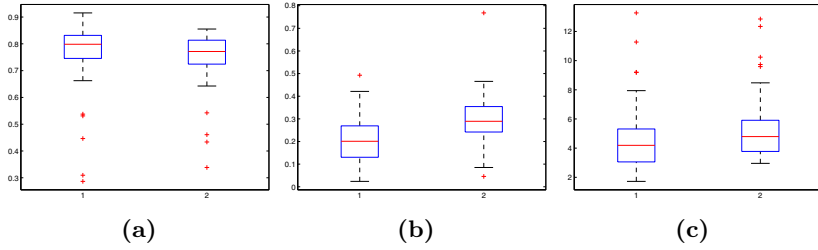
Fig. 1. Boxplots of the (a) Dice values, (b) false positive rates and MAD (c) for the joint framework (1) and the individual segmentation method (2)
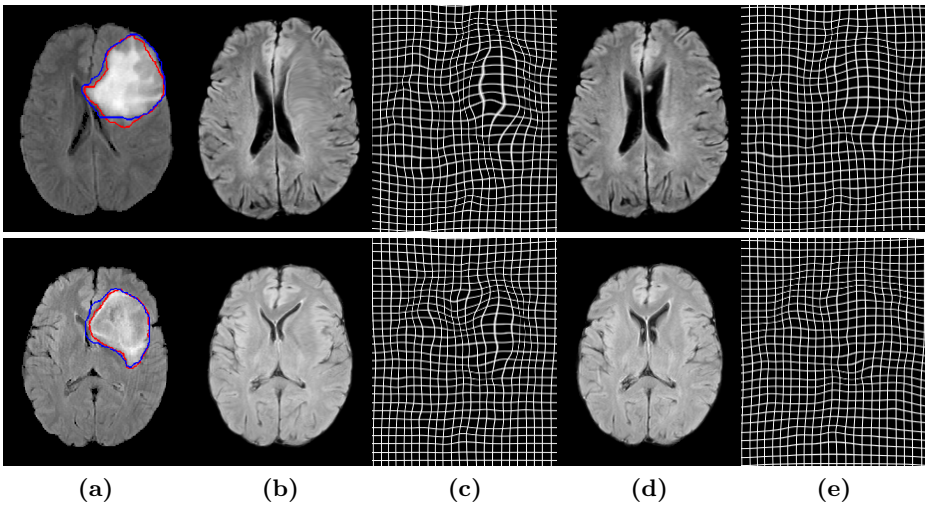
Fig. 2. Visual registration and segmentation results. (a) Segmented target image (manual (red) and automatic (blue) segmentations), (b) Individual registration, (d) our framework, (c,e) associated deformation fields.

4   Discussion

In this paper we have proposed a novel, efficient and principled method for combined tumor segmentation and dense registration. This was achieved through a two-layer interconnected graphical model that was optimized using a single shot approach towards optimal recovery of both variable spaces. Extensive validation concerning the case of low-grade gliomas was considered to evaluate the performance of the method, leading to very promising results. Introducing prior knowledge in the reference space as suggested in [2] is a straightforward extension of the proposed formulation which could further improve the performance of the method while allowing the characterization of tumors. Another possible extension of the method is the use of recent advances in MRF learning [16] towards

encoding prior knowledge using spatially varying coefficients of the corresponding segmentation framework that could be learned from training examples.

References

1. Kaus, M., Warfield, S., Nabavi, A., Black, P., Jolesz, F., Kikinis, R.: Automated Segmentation of MR Images of Brain Tumors. Radiology 218(2), 586–591 (2001)
2. Parisot, S., Duffau, H., Chemouny, S., Paragios, N.: Graph Based Spatial Position Mapping of Low-Grade Gliomas. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 508–515. Springer, Heidelberg (2011)
3. Friedman, J., Hastie, T., Tibshirani, R.: Additive Logistic Regression: a Statistical View of Boosting. The Annals of Statistics 38(2), 337–407 (2000)
4. Lee, C., Wang, S., Murtha, A., Brown, M., Greiner, R.: Segmenting Brain Tumors Using Pseudo–Conditional Random Fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 359–366. Springer, Heidelberg (2008)
5. Prastawa, M., Bullit, E., Moon, N., Leemput, K., Gerig, G.: Automatic Brain Tumor Segmentation by Subject Specific Modification of Atlas Priors. Academic Radiology 10(12), 1341–1348 (2003)
6. Stefanescu, R., Commowick, O., Malandain, G., Bondiau, P., Ayache, N., Pennec, X.: Non-rigid atlas to subject registration with pathologies for conformal brain radiotherapy. Springer (2004)
7. Bach Cuadra, M., De Craene, M., Duay, V., Macq, B., Pollo, C., Thiran, J.: Dense deformation field estimation for atlas-based segmentation of pathological MR brain images. Computer Methods and Programs in Biomedicine 84(2-3), 66–75 (2006)
8. Yezzi, A., Zollei, L., Kapur, T.: A variational framework for integrating segmentation and registration through active contours. Medical Image Analysis 7, 171–185 (2003)
9. Wyatt, P.P., Noble, J.A.: MAP MRF joint segmentation and registration of medical images. Medical Image Analysis 7(4), 539–552 (2003)
10. Gooya, A., Pohl, K.M., Bilello, M., Biros, G., Davatzikos, C.: Joint Segmentation and Deformable Registration of Brain Scans Guided by a Tumor Growth Model. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 532–540. Springer, Heidelberg (2011)
11. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
12. Boykov, Y., Funka-Lea, G.: Graph Cuts and Efficient ND Image Segmentation. International Journal of Computer Vision 70(2), 109–131 (2006)
13. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic MRFs: Setting the state of the art with primal-dual strategies. Computer Vision and Image Understanding 112(1), 14–29 (2008)
14. Michel, F., Bronstein, M., Bronstein, A., Paragios, N.: Boosted metric learning for 3D multi-modal deformable registration. In: ISBI, pp. 1209–1214. IEEE (2011)
15. Kang, S., Szeliski, R., Chai, J.: Handling occlusions in dense multi-view stereo. In: CVPR, vol. 1, pp. I–103. IEEE (2001)
16. Komodakis, N.: Efficient training for pairwise or higher order CRFs via dual decomposition. In: CVPR, pp. 1841–1848 (2011)

Registration Using Sparse Free-Form Deformations

Wenzhe Shi[1], Xiahai Zhuang[2], Luis Pizarro[1], Wenjia Bai[1], Haiyan Wang[1], Kai-Pin Tung[1], Philip Edwards[1], and Daniel Rueckert[1]

[1] Biomedical Image Analysis Group, Imperial College London, UK
[2] Shanghai Advanced Research Institute, Chinese Academy of Sciences, China

Abstract. Non-rigid image registration using free-form deformations (FFD) is a widely used technique in medical image registration. The balance between robustness and accuracy is controlled by the control point grid spacing and the amount of regularization. In this paper, we revisit the classic FFD registration approach and propose a sparse representation for FFDs using the principles of compressed sensing. The sparse free-form deformation model (SFFD) can capture fine local details such as motion discontinuities without sacrificing robustness. We demonstrate the capabilities of the proposed framework to accurately estimate smooth as well as discontinuous deformations in 2D and 3D image sequences. Compared to the classic FFD approach, a significant increase in registration accuracy can be observed in natural images (61%) as well as in cardiac MR images (53%) with discontinuous motions.

1 Introduction

The classic free-form deformation registration model (FFD) [1] is widely used for medical image registration. Several improvements of the method have been proposed including different optimization strategies [2, 3]. Despite its popularity, little effort has been devoted to improve the accuracy of the formulation compared to other registration methods such as optical flow [4–7] and the Demons approach [8, 9].

In this paper, we address one main difficulty of the classic FFD approach, namely the conflict between the robustness of the registration and the ability to model discontinuous deformations. This conflict stems from the fact that the FFD uses a smooth B-spline basis to model the contribution of each control point to the deformation. To model global and smooth deformations a coarse control point spacing is typically used. To allow for very localized deformations a fine control point spacing is required, making the method less robust. A conventional approach to address this issue uses a coarse-to-fine approach in which the initial coarse control point mesh is successively subdivided [1].

The standard smoothness constraints for different registration methods [1, 4, 8] assume that the transformation within a neighbourhood changes gradually since it is caused by a smooth motion. Combining the implicit smoothness of
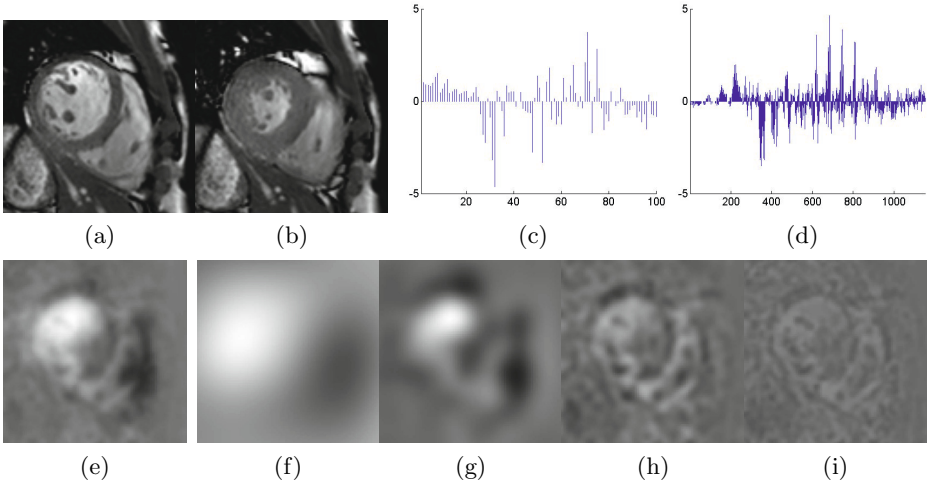
Fig. 1. (a) Target image; (b) source image; (c-d) coefficients of the B-spline basis for two levels (out of seven). Most coefficients are close to zero thus the sparsity assumption is valid; (e) deformation field induced by the FFD ϕ in x-direction; (f-i) FFD at four different levels.

the B-spline basis and explicit smoothness constraint in the regularization, the transformation model will produce smooth deformations.

As mentioned above, the control point grid spacing has a significant impact on the resulting transformation's ability to capture motion discontinuities robustly. Previous research focussed on the adaptive parametrization of the B-spline control point grid [10–12] driven by the intensity information in the images. An improved model should enable more control points to be placed in the area of the motion discontinuity.

Many other approaches to image registration have been proposed to overcome the conflict between robustness and discontinuity of the estimated motion in the field of optical flow [5–7]. Most recently, sparse coding methods have been proposed to evaluate the patch similarity between two images [13] and to constrain the transformation [14]. However there is no work ever reported to our best knowledge focused on motion discontinuity for the automatic FFD registration.

In this paper, we introduce a sparse representation for free-form deformations to estimate the transformation inspired by [13, 14]. This simple model uses the standard smoothness constraints and only imposes one assumption on the deformation, namely that the basis of the deformation is sparse in the parametric space. The assumption is generally true because the deformation between images is usually simpler than the actual images themselves. We use a multi-level FFD to represent the deformations in a parametric form. As can be seen from Figure 1, the sparsity assumption holds in general for multi-level FFDs with different control point spacings. Based on this assumption, we formulate the registration of two images using a sparse multi-level FFD representation of the control points.

We introduce a regularization term to impose smoothness at each level and a sparsity term to enforce coupled multi-level sparsity.

The novelty and contributions of this paper are the introduction of a sparsity model that avoids the *a priori* selection of an appropriate control point grid spacing. Furthermore, the approach reduces the conflict between global smoothness and local detail of the transformation by optimizing over the different FFD levels simultaneously with a sparsity constraint. These advantages allow for the robust estimation of deformation fields in the presence of discontinuous motion. We refer to this new approach as sparse free-form deformation model (SFFD). In the evaluation, we demonstrate that the proposed method can consistently capture localized motion with high accuracy.

2   Classic Free-Form Deformation Model

In the classic FFD registration [1], a non-rigid deformation $\phi = [X \ Y \ Z]^T$ is represented using a B-spline model in which the deformation is parameterized using a set of control points $\psi = [U \ V \ W]^T$ such that

$$\phi = \begin{bmatrix} B & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & B \end{bmatrix} \psi, \tag{1}$$

where B denotes the matrix of the B-spline basis functions. To find the optimal deformation between two images, the registration minimizes an energy functional E written as a function of ψ, which is typically a combination of two terms $E(\psi) := E_D(I_t, I_s \circ \phi) + E_R(\psi)$. The term E_D is a data constraint measuring the similarity between the target image I_t and the transformed source image $I_s \circ \phi$. The term E_R is a regularization constraint that enforces a smooth transformation. The energy function is typically minimized using gradient descent approaches [2] or discrete optimization approaches [15].

3   Sparse Free-Form Deformation Model

To be able to deal with large, global deformations and to improve the robustness, the classic FFD registration uses a multi-level approach: First, the optimal registration parameters are determined for a control point grid with large spacing. The grid is then successively subdivided to capture local deformations [1]. This requires an a-priori choice of the initial and final control point spacing. Furthermore, each level is optimized separately and once a level has been optimized it is no longer updated, leading to suboptimal registration results as can be seen in Figure 2. It was suggested in [14] that a realistic transformation can be easily embedded into a sparse representation. We postulate that an automatic selection of control points across different levels can be achieved by optimizing all FFD levels simultaneously while using a sparsity constraint.
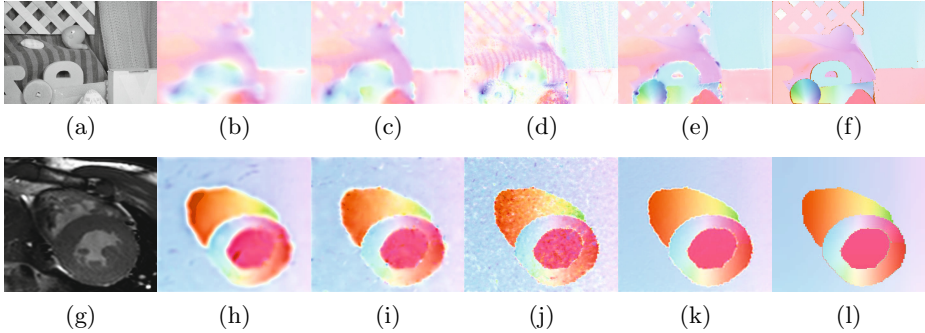
Fig. 2. Visual comparison between the classic FFD and the proposed SFFD using the colour scheme from [16]. The colour range corresponds to different direction and magnitude of the transformation. (a) and (g) show the rubber whale image from the Middlebury dataset and a frame from the cardiac B experiment, respectively;(f) and (l) display the ground truth transformations with noticeable motion discontinuities; (b-d) and (h-j) show the estimated motion with the classic FFD approach for 512/8mm, 256/4mm, and 64/1mm control point spacing; (e) and (k) exhibit the estimated motion with the proposed SFFD approach where $\lambda_S^N = 0.04$.

3.1   Sparse Free-Form Representation of Transformation

In this section, we propose to estimate the transformation ϕ with a sparse representation of the control points ψ. We use a m level FFD representation [10], $\psi = [U^1...U^m \ V^1...V^m \ W^1...W^m]^T$, as it is well suited for sparse representations. Accordingly, we utilize a multi-level B-splines basis $B = [B^1...B^m]$. The transformation ϕ is computed as in eq. (1) with the above redefinitions of ψ and B. As illustrated in Figure 1, a typical FFD is likely to be sparse in this representation.

Basis pursuit denoising [17] is a mathematical optimization problem that balances the trade-off between sparsity and reconstruction fidelity. In the context of image registration, the problem can be formulated as:

$$\arg\min_{\psi} \quad E(\psi) := ||I_t - I_s \circ \phi||_2^2 + ||\psi||_1\,, \tag{2}$$

where the first term corresponds to the sum of squared differences (SSD) between the target and the transformed source. The second term enforces sparsity of the solution ψ. In general, an arbitrary (dis)similarity measure can be utilised in the data term $E_D(I_t, I_s \circ \phi)$, including information theoretic measures such as mutual information (MI) or its normalized counterparts (NMI) [18].

Following these principles, we propose a novel registration approach, namely the sparse free-form deformation (SFFD) model as

$$\arg\min_{\psi} \quad E(\psi) := E_D(I_t, I_s \circ \phi) + \lambda_R \sum_{i\in[1,m]} E_R(\psi^i) + \lambda_S ||\psi||_1\,, \tag{3}$$

with constants $\lambda_R, \lambda_S \in \mathbb{R}^+$ weighting the regularization term and the sparsity term, respectively. Note that the regularization term imposes smoothness at each level of the multi-level FFD independently, while the sparsity term enforces coupled multi-level sparsity. This allows us to actively determine the importance of the control points across all levels in a joint manner, not independently as in the classic FFD framework. This strategy allows us to estimate deformations fields robustly and to preserve motion discontinuities, as it will be seen in the experimental validation. The SFFD model has the same computational complexity order as the classic FFD model.

We optimize eq. (3) using the interior point method of [19] that uses a log barrier function to make the sparsity term differentiable. The parameter λ_S is normalized between the data and the sparsity terms using the finite convergence to zero property. That is, for the L_1-regularized least squares problem, convergence is achieved for a finite value λ_{max} of λ_S. The value of λ_{max} can be determined using eq. (10) in [19]. In our experiments we use:

$$\lambda_S^N = \lambda_S/\lambda_{max} \,. \tag{4}$$

For completeness, the reader will find the partial derivatives of the similarity measures with respect to ψ in [20] for the SSD and in [2] for the NMI.

4   Results

4.1   Datasets

In this work, we have evaluated the proposed SFFD against the classic FFD model on four different datasets. The datasets we have used for evaluation include the Middlebury benchmarking dataset, 2D cardiac MR images with synthetic smooth (cardiac A) and discontinuous motion (cardiac B), and 3D cardiac MR image sequences (cardiac C).

For basic benchmarking we have used six pairs of 2D greyscale natural images from the Middlebury benchmark's training dataset [16]. The Middlebury benchmark's dataset contains deformations with multiple independently moving rigid objects and background. For the dataset the ground truth deformation between each pair of images is available.

In addition, we have tested our approach using 2D and 3D cardiac MR images. For the 2D cardiac MR images a synthetically generated transformation has been applied to the images using sin function as proposed in [21]. A single sin function is used to generate a group of 10 data with smooth motion using different magnitudes and frequencies. Also, a set of 10 discontinuous motions is generated where multiple sin functions are fused into a discontinuous motion using segmentation information which can be seen in Figure 2. The registration recovers the synthetic motion between the original image and the transformed image. For the above datasets, average error between the ground truth deformation and the deformation obtained after registration is measured.

We have also used 3D cardiac cine MR images from 13 normal volunteers to assess the proposed registration framework. The image resolution is $1.25 \times 1.25 \times 8mm$. We estimate the accuracy based on the tracking of myocardial boundaries of the left ventricle. Manual segmentation of endo- and epicardial surfaces are provided at both end-diastolic and end-systolic phases by a clinician. We then register the end-systolic (ES) phase directly to the end-diastolic (ED) phase. We evaluate surface distance between propagated myocardial surface and manual myocardial surface at end-systolic phase.

4.2   Implementation Details

During the optimization of the classic FFD, we use seven different levels of image resolution. The coarsest level has a size of around 64 voxels in each dimension. We use B-spline interpolation as evidence suggests that B-spline based interpolation is superior than linear interpolation [7, 2]. The coefficient for the smoothness penalty λ_R is set to 0.001 as in the original paper [1] for the NMI and 1 for the SSD due to the magnitude of the similarity metric.

One of most crucial parameters of the classic FFD registration is the control point grid spacing. We create a spacing at each image pyramid level by subdividing previous level's FFD. We have evaluated different initial control point spacings at the coarsest level varying from 512mm to 64mm where the final spacings at the final level varies from 8mm to 1mm. For the SFFD, we use a multi-level FFD with the coarsest level having a control point grid spacing of 64mm and finest level having a spacing of 1mm. The coarsest level should be reasonable according to data and the finest level should be around voxel size to embed a dense deformation. The number of image pyramid levels and the smoothness penalty are the same for both methods. Finally, NMI is used as similarity measures for the 3D cardiac MR images and SSD is used for the 2D cardiac MR images with simulated motion as well as for the Middlebury benchmark.

4.3   Evaluation

For the classic FFD, different spacing leads to significantly different results as showed in Table 1. Moreover, different datasets require different initial spacings to achieve best performances. It can be seen from the final column in Table 1 that the SFFD is robust against the choice of λ_S^N compare to the choice of spacing in FFD. There is little need to adjust λ_S^N across datasets to achieve near very good performance for individual data.

The median results using the multi-level FFD representation without sparsity constraint where $\lambda_S^N = 0$ are 0.68mm,0.025mm,0.072mm and 1.74mm for Middlebury and cardiac A,B,C datasets respectively. Thus all datasets benefit from the sparsity constraint with a consistent increase in registration accuracy. Moreover, the SFFD exhibits a significant improvement against the best results from the classic FFD where discontinuous motion presents. The improvement is most significant against ground-truth from the Middlebury dataset and from the 2D cardiac MR image dataset with synthetic discontinuous motion. An increasing

Table 1. Median of accuracies and standard deviation of median from the classic FFD with different initial and final spacing and from the SFFD with different normalized sparsity constraint λ_S^N. The units of the accuracy metrics are in the bracket. For the results, n/a means that the image is not large enough to contain the initial control point grid spacing, bold font means the best result and *means statistically significantly different from the best result of the classic FFD* using ttest with p-value 0.05. Final column is the median of the standard deviations of the results.

Initial/final spacing (FFD)	64/1 mm	128/2 mm	256/4 mm	512/8 mm	STD
Middleburry (mm)	0.69 ± 0.30	0.62 ± 0.27	$\mathbf{0.61 \pm 0.22}$	0.67 ± 0.33	0.108
Cardiac A (mm)	$.033 \pm .020^*$	$.017 \pm .010^*$	$.015 \pm .012$	$\mathbf{.012 \pm .010}$	0.014
Cardiac B (mm)	$.067 \pm .042^*$	$\mathbf{.045 \pm .016}$	$.048 \pm .011$	$.059 \pm .024^*$	0.034
Cardiac C (mm)	$\mathbf{1.68 \pm 0.28}$	$1.86 \pm 0.32^*$	$1.94 \pm 0.40^*$	n/a	0.619
$\lambda_S^N (SFFD)$	0.01	0.02	0.04	0.08	STD
Middleburry (mm)	$0.25 \pm 0.27^*$	$0.25 \pm 0.28^*$	$\mathbf{0.24 \pm 0.27^*}$	$0.25 \pm 0.28^*$	0.010
Cardiac A (mm)	$.010 \pm .006$	$.008 \pm .001$	$.007 \pm .001^*$	$\mathbf{.006 \pm .001^*}$	0.001
Cardiac B (mm)	$.023 \pm .005^*$	$.022 \pm .005^*$	$.021 \pm .005^*$	$\mathbf{.021 \pm .005^*}$	0.001
Cardiac C (mm)	1.57 ± 0.32	1.55 ± 0.31	$\mathbf{1.54 \pm 0.32}$	1.64 ± 0.32	0.091

ability to capture discontinuous motion while maintain robustness over smooth regions can be confirmed from a visual comparison in Figure 2.

Finally, in the 3D cardiac MR image sequences, we witness a limited improvement, 1.54 ± 0.32 *vs* 1.68 ± 0.28, in the result. Due to the lack of an objective ground truth, we measured the errors only on the LV myocardial surfaces, and hence it can only partially demonstrate the registration accuracy.

5   Conclusion

In this paper, we have developed a sparse free-form deformation model for registration which addresses some most important short-comings of the original FFD registration model. Control points across different FFD levels has been optimized simultaneously using a sparse representation. Compared to the classic FFD, the SFFD requires less parameter tuning across different datasets. The user no longer needs to choose an appropriate control point spacing a-priori. Our experiments have shown a consistent improvement compared to the original FFD aproach. The most significant improvement can be observed in Middleburry dataset (0.61 ± 0.22 *vs* 0.24 ± 0.27) and Cardiac B experiment (0.045 ± 0.016 *vs* 0.021 ± 0.005) where the deformation field exhibits both smooth and discontinuous motion.

References

1. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712–721 (1999)
2. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98(3), 278–284 (2010)

3. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.: Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29(1), 196–205 (2010)
4. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17(1-3), 185–203 (1981)
5. Mémin, E., Pérez, P.: Hierarchical estimation and segmentation of dense motion fields. International Journal of Computer Vision 46(2), 129–155 (2002)
6. Roth, S., Black, M.J.: On the spatial statistics of optical flow. In: ICCV 2005, vol. 1, pp. 42–49 (2005)
7. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2432–2439. IEEE (2010)
8. Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
9. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
10. Schnabel, J.A., Rueckert, D., Quist, M., Blackall, J.M., Castellano-Smith, A.D., Hartkens, T., Penney, G.P., Hall, W.A., Liu, H., Truwit, C.L., Gerritsen, F.A., Hill, D.L.G., Hawkes, D.J.: A Generic Framework for Non-rigid Registration Based on Non-uniform Multi-level Free-Form Deformations. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 573–581. Springer, Heidelberg (2001)
11. Rohlfing, T., Maurer Jr., C.R.: Intensity-Based Non-rigid Registration Using Adaptive Multilevel Free-Form Deformation with an Incompressibility Constraint. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 111–119. Springer, Heidelberg (2001)
12. Hansen, M.S., Larsen, R., Glocker, B., Navab, N.: Adaptive parametrization of multivariate b-splines for image registration. In: CVPR, pp. 1–8 (2008)
13. Roozgard, A., Barzigar, N., Cheng, S., Verma, P.: Dense image registration using sparse coding and belief propagation. In: International Conference on Signal Processing and Communication Systems, pp. 1–5 (2011)
14. Shen, X., Wu, Y.: Sparsity model for robust optical flow estimation at motion discontinuities. In: CVPR, pp. 2456–2463 (2010)
15. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. Medical Image Analysis 12(6), 731–741 (2008)
16. Baker, S., Scharstein, D., Lewis, J.P., Roth, S., Black, M.J., Szeliski, R.: A database and evaluation methodology for optical flow. In: ICCV, pp. 1–8 (2007)
17. Donoho, D.L., Huo, X.: Uncertainty principles and ideal atomic decomposition. IEEE Transactions on Information Theory 47(7), 2845–2862 (2001)
18. Studholme, C., Hill, D., Hawkes, D., et al.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognition 32(1), 71–86 (1999)
19. Kim, S.J., Koh, K., Lustig, M., Boyd, S., Gorinevsky, D.: An interior-point method for large-scale l1-regularized least squares. IEEE Journal of Selected Topics in Signal Processing 1(4), 606–617 (2007)
20. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Transactions on Image Processing 12(11), 1427–1442 (2003)
21. Pizarro, L., Delpiano, J., Aljabar, P., Ruiz-del-Solar, J., Rueckert, D.: Towards dense motion estimation in light and electron microscopy. In: ISBI: From Nano to Macro, pp. 1939–1942 (2011)

Registration of 3D Fetal Brain US and MRI

Maria Kuklisova-Murgasova[1], Amalia Cifor[1], Raffaele Napolitano[2],
Aris Papageorghiou[2], Gerardine Quaghebeur[3],
J. Alison Noble[1], and Julia A. Schnabel[1]

[1] Institute of Biomedical Engineering,
Department of Engineering Science, University of Oxford, U.K
[2] Nuffield Department of Obstetrics and Gynecology,
John Radcliffe Hospital, Oxford U.K
[3] Neuroradiology, John Radcliffe Hospital, Oxford U.K.

Abstract. We propose a novel method for registration of 3D fetal brain
ultrasound and a reconstructed magnetic resonance fetal brain volumes.
The reconstructed MR volume is first segmented using a probabilistic
atlas and an ultrasound-like image volume is simulated from the seg-
mentation of the MR image. This ultrasound-like image volume is then
affinely aligned with real ultrasound volumes of 27 fetal brains using a
robust block-matching approach which can deal with intensity artefacts
and missing features in ultrasound images. We show that this approach
results in good overlap of four small structures. The average of the co-
aligned US images shows good correlation with anatomy of the fetal brain
as seen in the MR reconstruction.

Keywords: Fetal 3D ultrasound, ultrasound and MR registration, fetal
ultrasound template, block matching.

1 Introduction

Fetal ultrasound (US) is the imaging modality of choice for assessing fetal de-
velopment in clinical practice. Recently, fetal brain magnetic resonance imaging
(MRI) has become an important modality for assessing fetal brain development
thanks to development of fast-spin-echo sequences for acquisition of 2D slices
which can deal with unpredictable and fast fetal motion. Emerging volumetric
fetal brain MRI reconstruction and segmentation techniques can now support
quantitative studies of fetal brain development [1]. Recent advances in US tech-
nology allow 3D acquisitions of the fetal brain which can potentially facilitate
volumetric assessment of fetal brain. US, compared to MRI, has a well demon-
strated safety, a lower cost and causes less stress to the patients. Models of fetal
brain anatomy constructed from fetal brain MRI can facilitate development of
image analysis methods for fetal 3D ultrasound and their validation.

In this we paper we propose a method for 3D alignment of fetal brain MRI
to US, which is one of the important tools with potential to facilitate automatic
image analysis of 3D brain fetal US. Registration of brain MRI and US was

previously proposed for image guided neurosurgery in adults to non-rigidly correct for brain-shift during surgery [2] as well as for rigid alignment of adult or neonatal brain [3]. The major challenge in aligning MR and US images is that there does not exist a simple intensity mapping between the two modalities, as they visualize complementary features. While MRI offers good contrast between soft tissues, especially white matter (WM), grey matter (GM) and cerebro-spinal fluid (CSF), US depicts surfaces of the structures with superior spatial resolution compared to MRI, while the WM-GM boundary does not appear in US at all. Roche at al. [3] suggested to estimate a non-linear relationship between MR and US image intensities and used a generalized correlation ratio as a similarity measure. Arbel et al. [2] proposed to perform segmentation of the brain MRI followed by simulation of an US-like image which is then registered with the US image using local normalized cross-correlation (NCC) as a similarity measure. As the relationship of MR and US intensities in fetal brain is difficult to describe, we propose to simulate US-like image from the segmented MRI, similar to [2]. Features of fetal brain visible in 3D US in relation to the anatomy have been described in the clinical literature [4]. We can therefore build on this knowledge and convert a segmentation of the anatomical structures in fetal brain MRI to into features which appear in fetal US (Sec. 2). US images are corrupted by intensity artefacts, such as shadows and attenuation. Compared to cranial sonography in adults and neonates, where radiographer can position the probe next to the opening in the skull, such artefacts are significantly more pronounced in US of the fetal head, due to difficulties with positioning of the probe and interference with maternal tissues. Additionally, partly calcified and not completely fused fetal skull causes uneven loss of signal strength. Wein et.al. [5] suggested to estimate the attenuation from knowledge of physical properties of the scanned organ, however, these are not well described for fetal head and brain. Therefore in this paper we propose to apply a robust block-matching strategy [6] with local CC as a similarity measure (Sec. 3) to overcome the limitations of a global similarity measure [3] and gradient descent optimisation [2] in presence of significant intensity artefacts in fetal brain US. The method has been applied to align 27 3D US images at gestational ages (GA) 18 to 22 weeks with an MRI of a single subject of GA 23 weeks (Sec. 4). In spite of a relatively significant GA difference between subjects we achieved good volume overlaps for four small structures. The estimated transformations were used to coalign the US images with the MRI template. The average of the coaligned US images depicts the near-complete anatomy of the fetal brain as visible in 3D US.

2   Construction of US-Like Template Image from MRI

2.1   Reconstruction of Fetal Brain MRI

We constructed the US-like template from MRI of a single fetus. As US depicts structures of small scale compared to spatial resolution of MRI (e.g. septi pellucidi), it was essential to perform a high quality reconstruction of the MRI. To achieve this, the fetal brain volume was reconstructed from stacks of thin
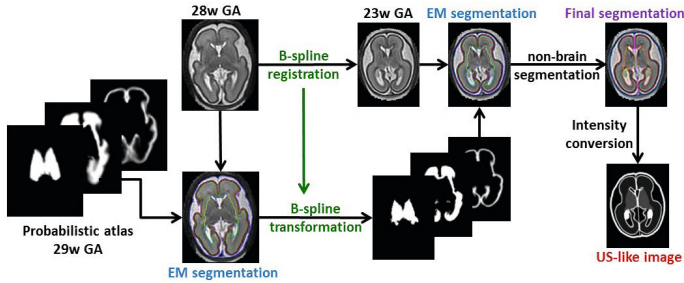
Fig. 1. Overview of the segmentation of the MR template and conversion into US-like image

slices (2.5mm) using a super-resolution approach [7]. High signal to noise ratio was achieved by carefully choosing the point-spread-function (PSF) to match the real acquisition. The PSF was approximated by a 3D Gaussian with full width at half maximum (FWHM) equal to slice thickness in through-plane direction to approximate the slice selection profile (truncated sinc function) and 1.2×resolution for in-plane direction to approximate the sinc function. The reconstructed volume is presented in Fig. 3a.

2.2 Segmentation of Brain Structures in MRI

The reconstructed MR volume of a single fetus with GA 23 weeks was segmented and converted to the US-like image, see overview in Fig. 1. First, we performed an automatic segmention using the EM algorithm [8] into six brain structures - WM, cortex, CSF, deep grey matter (DGM), brainstem and cerebellum. As there are no fetal probabilistic atlases publicly available at present, we used a spatio-temporal neonatal probabilistic atlas [9]. The youngest time-point in this atlas is 29 weeks GA. The difference in shape and cortical folding for these GA requires a flexible non-rigid registration and for this we used non-rigid B-spline registration with final control point spacing 2.5mm and normalized mutual information. We found that registration of a blurry probabilistic template with such resolution is unstable, but subject-to-subject registration across such age-range produces good results for the kind of structures we are interested in. We therefore first segmented the brain MR volume of another fetus with GA 28 weeks. This image was then registered with the MR template (23 weeks GA) The segmentation of the older fetus was then used as a prior during EM segmentation of the younger. The cortex prior was slightly blurred before segmentation to account for cortical folding of the older subject. Though the cortex-WM separation still had some errors this was not an issue as their boundary is not visible in US and these two tissues were joined together. At this gestation, the cerebellum does not appear as one homogeneous structure in US, but high fluid content of cerebellar hemispheres gives it appearance of two dark cysts surrounded by bright tissue [10], see Fig. 3. The cerebellar cysts were therefore manually excluded from the automatic segmentation of the cerebellum.

2.3   Segmentation of Non-brain Structures in MRI

Some of the most echogenic landmarks in fetal US depict the non-brain structures, especially the skull, choroid plexus, septi pellucidi and membrane of the falx in brain midline (see Fig. 3). We therefore performed semi-automatic or manual segmentation of these four structures in the MR template.

Segmentation of the skull in MRI is a difficult task as it appears dark and often borders other tissues also appearing dark on T2w MRI. Additionally, the fetal skull is rather thin compared to the resolution during acquisition, making it difficult to delineate due to the partial volume effect. The skull was therefore estimated from the brain mask created by joining automatic segmentations of five brain tissues and CSF. A distance transform from this brain mask was calculated, and it was visually determined that the voxels with distance up to 2mm can be labeled as the skull. The segmentation of the skull was then manually corrected.

In fetal brain US, the brain tissue (with exception of the corpus callosum) is clearly divided into a right and left hemisphere. This however is only partially visible in MRI due to insufficient resolution and partial volume effects along the midline. We therefore artificially separated the WM segmentation into two parts (by removing the voxels closer than 1.5mm to the WM surface) and calculated distance transforms from these two WM cores. All the brain structures (except for WM) were separated into right or left by removing voxels with the roughly same distance from the two WM cores (the midline voxels). The midline was then added as another structure to simulate the falx visible on fetal brain US. Additional two structures which are significant landmarks on fetal brain US - choroid plexus and septum pellucidum - were segmented manually.

The segmentation pipeline described here included manual steps, however this template will be used for inter-subject registration and therefore does not need to be repeatedly segmented. Additionally, it can be used for atlas-based segmentation of MRI of new subjects.

2.4   Converting the Segmentation into US-Like Image

US B-mode images are created by reflections of tissue surfaces with difference in acoustic impedance and speckle patterns produced by interference of tissue microstructure with the sound waves. These intensity patterns are further affected by intensity attenuation (or signal loss along the beam direction) and shadows, which occur when beam is fully reflected by a strong reflector. In this work we assume that the fetal brain US is mainly composed of echogenicity of the tissues and neglect the reflections at the tissue boundaries and intensity artefacts. The visibility of brain surface is also due to a presence of a highly echogenic thin tissue layer [4] not visible in MRI. We therefore convert the segmentation to artefact-free US-like image in which each region of interest is assigned uniform intensity representing the average echogenicity of this region. We did not include a model of speckle in the US-like image, as it is not useful for guiding inter-subject alignment. The US-like image is then registered with real US images using a method robust to the artefacts and missing features (Sec. 3).

The segmented anatomical structures were assigned a visually determined intensity value in order from brightest to darkest: 1. skull, 2. choroid plexus, septi pellucidi and midline, 3. brain surface, 4. cerebellum, 5. DGM and brainstem, 6. cerebral hemispheres. As we use local NCC as similarity measure, the exact value of tissue intensity is not important, but the correctness of the order of brightness is essential. The US-like image is shown in Fig. 2c.

3   Alignment of MRI and US

The block-matching algorithm [6] iteratively estimates the alignment T between two images by alternating between two stages. In the first stage, for each block in the source image b_S^i, the most similar (homologous) block b_T^j is found in the corresponding exploration neighborhood (determined by the latest estimate of the transformation T) in the target image. We use NCC to find the most similar blocks, as the construction of US-like image from MRI effectively changes our registration problem from multi-modal to mono-modal. NCC is calculated independently for each pair of blocks, ensuring low sensitivity to intensity artefacts. The centroids of the pairs of the homologous blocks define a displacement field which is regularized in the second stage. All the blocks with variance smaller than a pre-defined threshold are excluded in the first stage so that only blocks carrying features are used for estimation of the transformation T, thus improving robustness towards missing features.

In the second stage, the rigid or affine transformation is estimated from the displacement field using least trimmed squared regression (LTS) [11]. The LTS estimator reduces the influence of outliers by minimizing the sum of a given number (h) of smallest squared residuals. A residual error is obtained as the difference between the displacement d_i at centroid C_S^i and the one obtained by applying the estimated transformation to it. Such robust estimation of transformation is essential to remove influence of displacements for the source blocks which have no corresponding target block due to missing features in US images.

$$T^* = argmin \sum_1^h \| (C_S^i + d(i)) - T[C_S^i] \|^2$$

4   Results

The proposed registration method was applied to 27 US volumes of the fetal head, with GA 18-22 weeks. Images were acquired with a Philips HD9 or iU22 machine using a 3D transabdominal probe with mechanical steer. The images were first consistently re-oriented and scaled to a similar size using a scaling factor derived from the age of the subject. The block-matching algorithm was then used to align the US-like image derived from the MR template with each US image. We determined the rigid alignment first followed by the affine alignment. The block-matching was applied in two resolution levels (isotropic voxel size
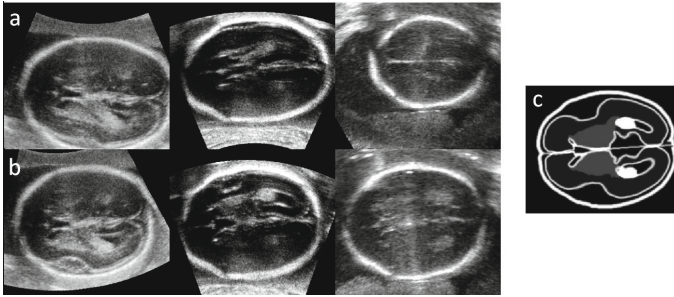
Fig. 2. A 2D view of 3D US of three fetuses before (a) and after (b) alignment with the MR template. After alignment the view shows transventricular plane in all three cases. Compare to US-like image (c). Note that the difference in cortical folding in the area around hippocampus stems from differences in GA.

$2mm$ and $1mm$). We used blocks of $3 \times 3 \times 3$ voxels and neighbourhoods with 7 voxels in each direction. The variance threshold used to exclude blocks was $0.04\times$ intensity range. The LTS transformation was estimated using 75% of the displacements. Figs. 2a,b show US images before and after alignment with the US-like image (Fig. 2c).

In each of the 27 ultrasound images, four small structures - choroid plexus (CP), cavum septi pellucidi (CSP), posterior ventricular cavity (VC) and cerebellar hemisphere (CH) - were manually delineated by a clinical expert. All these structures were also delineated in the MR template. The average volume (Dice) overlaps for these four structures are presented in Table 1 (fourth row). We achieved average volume overlaps around 0.5 which we consider very good, as the residual differences in brain shape between subjects and ages are still expected to be significant after global affine alignment. In all cases, all four structures were overlapping after alignment and therefore none of the registrations failed. To further illustrate the quality of alignment, Fig. 3c presents the average of all 27 US volumes aligned using the inverse of the estimated transformations.

To demonstrate that converting MRI to US-like image is essential we compared our proposed method with two recognized approaches which can perform multimodal registration of original MRI with US images: 1. Registration using normalized mutual information using IRTK[1]; 2. Block-matching method using correlation ratio (CR). In the second method the similarity measure had to be changed compared to our proposed approach because NCC is not suitable for multimodal registration and CR is a suitable multimodal similarity measure to be used in conjunction with block-matching [6]. CR can be implemented as a statistical measure or the functional relationship between intensities of the target and source image can be defined by a parametric model. Due to the small size of the blocks in our approach we consider affine relationship of the intensities a reasonable assumption. In this case CR is equivalent to squared NCC. In the first multimodal approach the average volume overlaps were much smaller

---

[1] http://www.doc.ic.ac.uk/~dr/software/

Table 1. Average volume overlaps and their standard deviations after registration of the source image (second column) to US images. The last column shows number of failed cases.

Registration	Similarity	Source	CP	VC	CH	CSP	failed
Gradient descent	NMI	MRI	0.18±0.24	0.12±0.20	0.12±0.16	0.13±0.16	14
block-matching	CR	MRI	0.47±0.15	0.41±0.18	0.48±0.16	0.37±0.15	0
block-matching	CR	US-like	0.56±0.12	0.39±0.15	0.54±0.13	0.40±0.10	0
block-matching	NCC	US-like	0.57±0.10	0.42±0.14	0.56±0.09	0.43±0.10	0
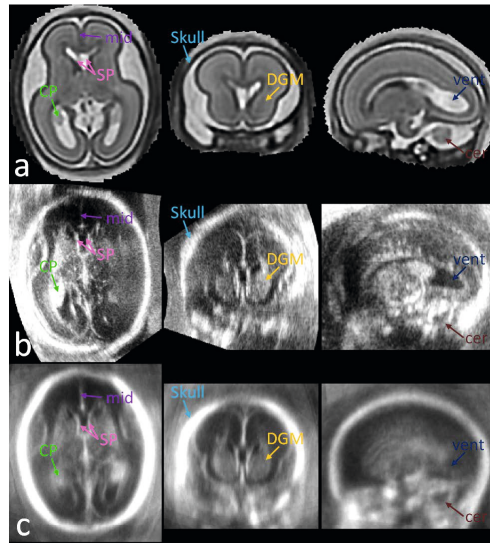
Fig. 3. Axial, coronal and sagittal view of (a) brain MRI of a single fetus used as a template; (b) brain 3D US of a single fetus after alignment with MR template; (c) average of the 27 fetal brain 3D US aligned with the MR template. Highlighted structures: Choroid plexus (CP), septi pellucidi (SP), falx (mid), deep grey matter (DGM), cerebellum (cer), posterior ventricular cavity (vent), skull.

and in a number of the cases the registration failed on the criterion that at least three out of four segmented structures have non-zero overlap (see Table 1, first row). In the multimodal block-matching approach none of the registrations failed but average overlaps were significantly smaller (see Table 1, second row). This demonstrates that US-like image offers better correspondences with US images than original MRI. The third experiments demonstrates, that the more constrained NCC yields better performance for the block-matching of the US-like image with real US images than the less constrained CR (see Table 1, third row).

5   Discussion and Conclusion

In this paper we have presented a novel method for affine registration of fetal brain MR and US volumes. The method succeeded in all 27 cases and we achieved good volume overlaps for four small structures. The average of the co-aligned US volumes revealed near-complete anatomy of the fetal brain. The results suggest that 3D US in conjunction with MR prior has potential for development of new medical image analysis tools for assessment fetal brain development.

References

1. Rajagopalan, V., Scott, J., Habas, P.A., Kim, K., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C.: Spatiotemporal Morphometry of Adjacent Tissue Layers with Application to the Study of Sulcal Formation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part II. LNCS, vol. 6892, pp. 476–483. Springer, Heidelberg (2011)
2. Arbel, T., Morandi, X., Comeau, R.M., Louis Collins, D.: Automatic Non-linear MRI-Ultrasound Registration for the Correction of Intra-operative Brain Deformations. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 913–922. Springer, Heidelberg (2001)
3. Roche, A., Pennec, X., Malandain, G., Ayache, N.: Rigid registration of 3D ultrasound with MR images: A new approach combining intensity and gradient information. IEEE Transactions on Medical Imaging 20(10), 1038–1049 (2001)
4. Monteagudo, A., Timor-Tritsch, I.E.: Normal sonographic development of the central nervous system from the second trimester onwards using 2D, 3D and transvaginal sonography. Prenatal Diagnosis 29(4), 326–339 (2009)
5. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)
6. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. Image and Vision Computing 19(12), 25–31 (2001)
7. Murgasova, M., Quaghebeur, G., Hajnal, J.V., Schnabel, J.A.: Unified framework for superresolution reconstruction of 3D fetal brain mr images from 2D slices with intensity correction and outlier rejection. In: MICCAI Workshop on Image Analysis of Human Brain Development, IAHBD (2011)
8. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE TMI 18(10), 885–896 (1999)
9. Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S.J., Doria, V., Serag, A., Gousias, I.S., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D.: A dynamic 4D probabilistic atlas of the developing brain. NeuroImage 54(4), 2750–2763 (2011)
10. Hashimoto, K., Shimizu, T., Shimoya, K., Kanzaki, T., Clapp, J.F., Murata, Y.: Fetal cerebellum: US appearance with advancing gestational age. Radiology 221(1), 70–74 (2001)
11. Rousseeuw, P.J.: Least median of squares regression. Journal of the American Statistical Association 79(388), 871–880 (1984)

Author Index